

Transcription

CHAPTER OUTLINE

1 Introduction

2 Transcription Occurs by Base Pairing in a "Bubble" of Unpaired DNA

- RNA polymerase separates the two strands of DNA in a transient "bubble" and uses one strand as a template to direct synthesis of a complementary sequence of RNA.
- The length of the bubble is ~12 to 14 bp, and the length of RNA-DNA hybrid within it is ~8 to 9 bp.

3 The Transcription Reaction Has Three Stages

- RNA polymerase initiates transcription after binding to a promoter site on DNA.
- During elongation the transcription bubble moves along DNA and the RNA chain is extended in the 5'–3' direction.
- When transcription stops, the DNA duplex reforms and RNA polymerase dissociates at a terminator site.

4 Phage T7 RNA Polymerase Is a Useful Model System

- T3 and T7 phage RNA polymerases are single polypeptides with minimal activities in recognizing a small number of phage promoters.
- Crystal structures of T7 RNA polymerase with DNA identify the DNA-binding region and the active site.

5 A Model for Enzyme Movement Is Suggested by the Crystal Structure

- DNA moves through a groove in yeast RNA polymerase that makes a sharp turn at the active site.
- A protein bridge changes conformation to control the entry of nucleotides to the active site.

6 Bacterial RNA Polymerase Consists of Multiple Subunits

- Bacterial RNA core polymerases are ~500 kD multisubunit complexes with the general structure $\alpha_2\beta\beta'$.
- DNA is bound in a channel and is contacted by both the β and β' subunits.

7 RNA Polymerase Consists of the Core Enzyme and Sigma Factor

- Bacterial RNA polymerase can be divided into the $\alpha_2\beta\beta'$ core enzyme that catalyzes transcription and the sigma subunit that is required only for initiation.
- Sigma factor changes the DNA-binding properties of RNA polymerase so that its affinity for general DNA is reduced and its affinity for promoters is increased.

- Binding constants of RNA polymerase for different promoters vary over six orders of magnitude, corresponding to the frequency with which transcription is initiated at each promoter.

11.8 The Association with Sigma Factor Changes at Initiation

- When RNA polymerase binds to a promoter, it separates the DNA strands to form a transcription bubble and incorporates up to nine nucleotides into RNA.
- There may be a cycle of abortive initiations before the enzyme moves to the next phase.
- Sigma factor may be released from RNA polymerase when the nascent RNA chain reaches eight to nine bases in length.

11.9 A Stalled RNA Polymerase Can Restart

- An arrested RNA polymerase can restart transcription by cleaving the RNA transcript to generate a new 3' end.

11.10 How Does RNA Polymerase Find Promoter Sequences?

- The rate at which RNA polymerase binds to promoters is too fast to be accounted for by random diffusion.
- RNA polymerase probably binds to random sites on DNA and exchanges them with other sequences very rapidly until a promoter is found.

11.11 Sigma Factor Controls Binding to DNA

- A change in association between sigma factor and holoenzyme changes binding affinity for DNA so that core enzyme can move along DNA.

11.12 Promoter Recognition Depends on Consensus Sequences

- A promoter is defined by the presence of short consensus sequences at specific locations.
- The promoter consensus sequences consist of a purine at the startpoint, the hexamer TATAAT centered at ~10, and another hexamer centered at ~35.
- Individual promoters usually differ from the consensus at one or more positions.

11.13 Promoter Efficiencies Can Be Increased or Decreased by Mutation

- Down mutations to decrease promoter efficiency usually decrease conformance to the consensus sequences, whereas up mutations have the opposite effect.
- Mutations in the ~35 sequence usually affect initial binding of RNA polymerase.

- Mutations in the ~10 sequence usually affect the melting reaction that converts a closed to an open complex.

11.14 RNA Polymerase Binds to One Face of DNA

- The consensus sequences at ~35 and ~10 provide most of the contact points for RNA polymerase in the promoter.
- The points of contact lie on one face of the DNA.

11.15 Supercoiling Is an Important Feature of Transcription

- Negative supercoiling increases the efficiency of some promoters by assisting the melting reaction.
- Transcription generates positive supercoils ahead of the enzyme and negative supercoils behind it, and these must be removed by gyrase and topoisomerase.

11.16 Substitution of Sigma Factors May Control Initiation

- *E. coli* has several sigma factors, each of which causes RNA polymerase to initiate at a set of promoters defined by specific ~35 and ~10 sequences.
- σ^{70} is used for general transcription, and the other sigma factors are activated by special conditions.

11.17 Sigma Factors Directly Contact DNA

- σ^{70} changes its structure to release its DNA-binding regions when it associates with core enzyme.
- σ^{70} binds both the ~35 and ~10 sequences.

11.18 Sigma Factors May Be Organized into Cascades

- A cascade of sigma factors is created when one sigma factor is required to transcribe the gene coding for the next sigma factor.
- The early genes of phage SP01 are transcribed by host RNA polymerase.
- One of the early genes codes for a sigma factor that causes RNA polymerase to transcribe the middle genes.
- Two of the middle genes code for subunits of a sigma factor that causes RNA polymerase to transcribe the late genes.

11.19 Sporulation Is Controlled by Sigma Factors

- Sporulation divides a bacterium into a mother cell that is lysed and a spore that is released.

- Each compartment advances to the next stage of development by synthesizing a new sigma factor that displaces the previous sigma factor.
- Communication between the two compartments coordinates the timing of sigma factor substitutions.

11.20 Bacterial RNA Polymerase Terminates at Discrete Sites

- Termination may require both recognition of the terminator sequence in DNA and the formation of a hairpin structure in the RNA product.

11.21 There Are Two Types of Terminators in *E. coli*

- Intrinsic terminators consist of a G-C-rich hairpin in the RNA product followed by a U-rich region in which termination occurs.

11.22 How Does Rho Factor Work?

- Rho factor is a terminator protein that binds to a *rut* site on nascent RNA and tracks along the RNA to release it from the RNA-DNA hybrid structure at the RNA polymerase.

11.23 Antitermination Is a Regulatory Event

- Termination is prevented when antitermination proteins act on RNA polymerase to cause it to read through a specific terminator or terminators.
- Phage lambda has two antitermination proteins, pN and pQ, that act on different transcription units.

11.24 Antitermination Requires Sites That Are Independent of the Terminators

- The site where an antiterminator protein acts is upstream of the terminator site in the transcription unit.
- The location of the antiterminator site varies in different cases and can be in the promoter or within the transcription unit.

11.25 Termination and Antitermination Factors Interact with RNA Polymerase

- Several bacterial proteins are required for lambda pN to interact with RNA polymerase.
- These proteins are also involved in antitermination in the *rnm* operons of the host bacterium.
- The lambda antiterminator pQ has a different mode of interaction that involves binding to DNA at the promoter.

11.26 Summary

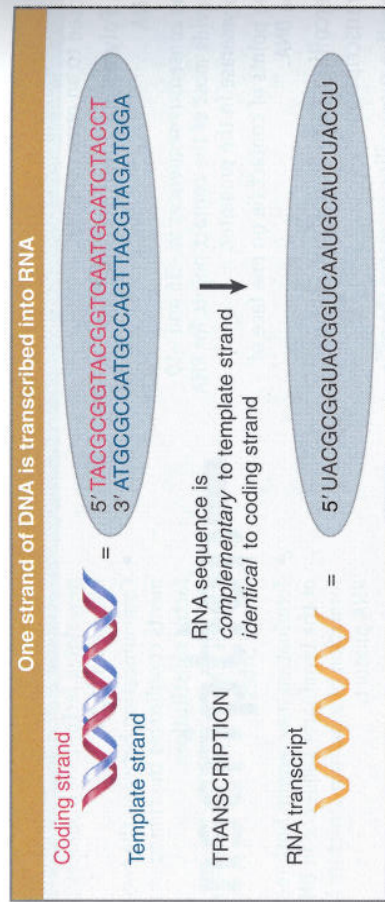


FIGURE 11.1 The function of RNA polymerase is to copy one strand of duplex DNA into RNA.

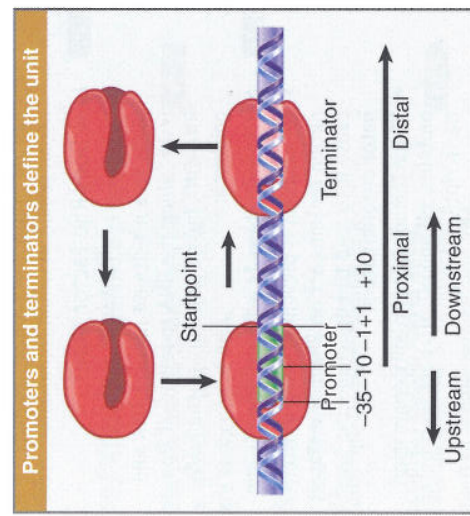


FIGURE 11.2 A transcription unit is a sequence of DNA transcribed into a single RNA, starting at the promoter and ending at the terminator.

11.1 Introduction

Transcription involves synthesis of an RNA chain representing one strand of a DNA duplex. When we say “representing,” we mean that the RNA is *identical in sequence* with one strand of the DNA, which is called the **coding strand**. It is *complementary* to the other strand, which provides the **template strand** for its synthesis. **FIGURE 11.1** recapitulates the relationship between double-stranded DNA and its single-stranded RNA transcript.

RNA synthesis is catalyzed by the enzyme **RNA polymerase**. Transcription starts when RNA polymerase binds to a special region, the **promoter**, at the start of the gene. The promoter surrounds the first base pair that is transcribed into RNA, the **start point**. From this point, RNA polymerase moves along the template, synthesizing RNA, until it reaches a **terminator** (*t*) sequence. This action defines a **transcription unit** that extends from the pro-

transcription (or other stages of gene expression) are sometimes regulated.

Within this context, there are two basic questions in gene expression:

- How does RNA polymerase find promoters on DNA? This is a particular example of a more general question: How do proteins distinguish their specific binding sites in DNA from other sequences?
- How do regulatory proteins interact with RNA polymerase (and with one another) to activate or to repress specific steps in the initiation, elongation, or termination of transcription?

In this chapter, we analyze the interactions of bacterial RNA polymerase with DNA from its initial contact with a gene, through the act of transcription, and then finally its release when the transcript has been completed. Chapter 12, The Operon, describes the various means by which regulatory proteins can assist or prevent bacterial RNA polymerase from recognizing a particular gene for transcription. Chapter 13, Regulatory RNA, discusses other means of regulation, including the use of small RNAs, and considers how these interactions can be connected into larger regulatory networks. In Chapter 14, Phage Strategies, we consider how individual regulatory interactions can be connected into more complex networks. In Chapter 24, Promoters and Enhancers, and Chapter 25, Activating Transcription, we consider the analogous reactions between eukaryotic RNA polymerases and their templates.

11.2 Transcription Occurs by Base Pairing in a “Bubble” of Unpaired DNA

Key concepts

- RNA polymerase separates the two strands of DNA in a transient “bubble” and uses one strand as a template to direct synthesis of a complementary sequence of RNA.
- The length of the bubble is ~12 to 14 bp, and the length of RNA-DNA hybrid within it is ~8 to 9 bp.

Transcription takes place by the usual process of complementary base pairing. **FIGURE 11.3** illustrates the general principle of transcription. RNA synthesis takes place within a “transcription bubble,” in which DNA is transiently separated

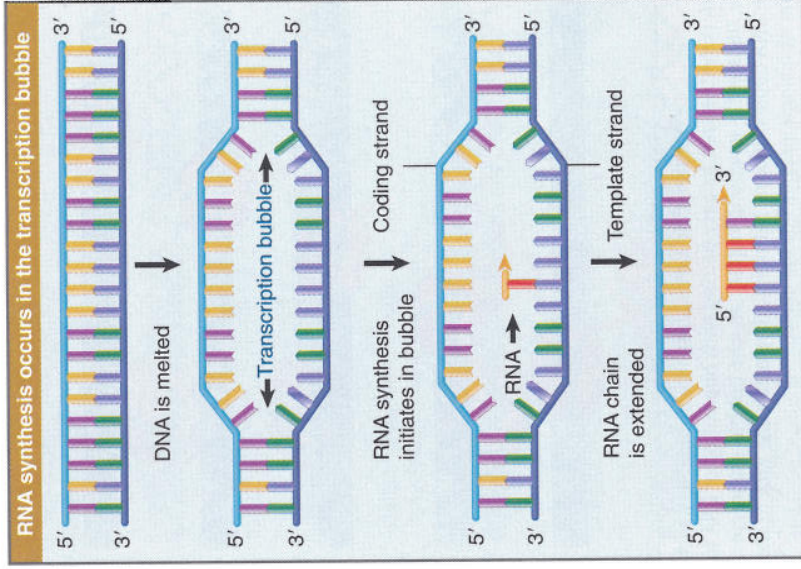


FIGURE 11.3 DNA strands separate to form a transcription bubble. RNA is synthesized by complementary base pairing with one of the DNA strands.

into its single strands and the template strand is used to direct synthesis of the RNA strand.

The RNA chain is synthesized from the 5' end toward the 3' end. The 3'-OH group of the last nucleotide added to the chain reacts with an incoming nucleoside 5' triphosphate. The incoming nucleoside loses its terminal two phosphate groups (γ and β); its α group is used in the phosphodiester bond linking it to the chain. The overall reaction rate is ~40 nucleotides/second at 37° C (for the bacterial RNA polymerase); this is about the same as the rate of translation (15 amino acids/sec), but much slower than the rate of DNA replication (800 bp/sec).

RNA polymerase creates the transcription bubble when it binds to a promoter. **FIGURE 11.4** shows that as RNA polymerase moves along the DNA, the bubble moves with it and the RNA chain grows longer. The process of base pairing and base addition within the bubble is catalyzed and scrutinized by the enzyme.

The structure of the bubble within RNA polymerase is shown in the expanded view of **FIGURE 11.5**. As RNA polymerase moves along the DNA template, it unwinds the duplex at the

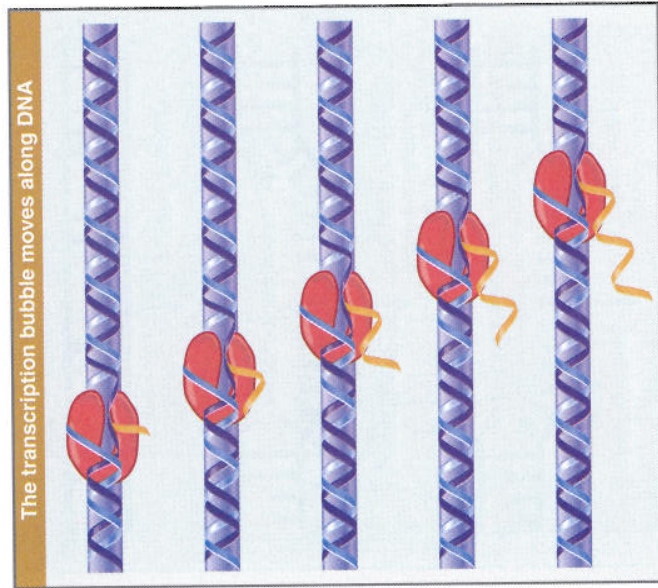


FIGURE 11.4 Transcription takes place in a bubble, in which RNA is synthesized by base pairing with one strand of DNA in the transiently unwound region. As the bubble progresses, the DNA duplex reforms behind it, displacing the RNA in the form of a single polynucleotide chain.

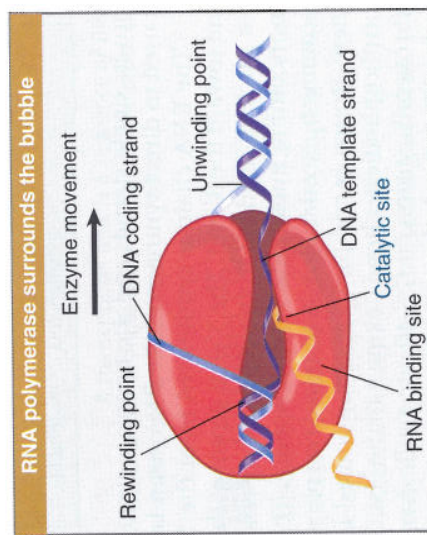


FIGURE 11.5 During transcription, the bubble is maintained within bacterial RNA polymerase, which unwinds and rewinds DNA and synthesizes RNA.

front of the bubble (the unwinding point), and rewinds the DNA at the back (the rewinding point). The length of the transcription bubble is ~12 to 14 bp, but the length of the RNA-DNA hybrid region within it is shorter.

There is a major change in the topology of DNA extending over ~1 turn, but it is not clear how much of this region is actually base paired

with RNA at any given moment. Certainly the RNA-DNA hybrid is short and transient. As the enzyme moves on, the DNA duplex reforms, and the RNA is displaced as a free polynucleotide chain. Roughly the last twenty-five ribonucleotides added to a growing chain are complexed with DNA and/or enzyme at any moment.

11.3 The Transcription Reaction Has Three Stages

Key concepts

- RNA polymerase initiates transcription after binding to a promoter site on DNA.
- During elongation the transcription bubble moves along DNA and the RNA chain is extended in the 5'–3' direction.
- When transcription stops, the DNA duplex reforms and RNA polymerase dissociates at a terminator site.

The transcription reaction can be divided into the stages illustrated in **FIGURE 11.6**, in which a bubble is created, RNA synthesis begins, the bubble moves along the DNA, and finally the bubble is terminated:

- **Template recognition** begins with the binding of RNA polymerase to the double-stranded DNA at a promoter to form a “closed complex.” The strands of DNA are then separated to form the “open complex” that makes the template strand available for base pairing with ribonucleotides. The transcription bubble is created by a local unwinding that begins at the site bound by RNA polymerase.
- **Initiation** describes the synthesis of the first nucleotide bonds in RNA. The enzyme remains at the promoter while it synthesizes the first ~9 nucleotide bonds. The initiation phase is protracted by the occurrence of abortive events, in which the enzyme makes short transcripts, releases them, and then starts synthesis of RNA again. The initiation phase ends when the enzyme succeeds in extending the chain and clears the promoter. *The sequence of DNA needed for RNA polymerase to bind to the template and accomplish the initiation reaction defines the promoter.* Abortive initiation probably involves synthesizing an RNA chain that

fills the active site. If the RNA is released, the initiation is aborted and must start again. Initiation is accomplished if and when the enzyme manages to move along the template to move the next region of the DNA into the active site.

- During **elongation** the enzyme moves along the DNA and extends the growing RNA chain. As the enzyme moves, it unwinds the DNA helix to expose a new segment of the template in single-stranded condition. Nucleotides are covalently added to the 3' end of the growing RNA chain, forming an RNA-DNA hybrid in the unwound region. Behind the unwound region, the DNA template strand pairs with its original partner to reform the double helix. The RNA emerges as a free single strand. *Elongation involves the movement of the transcription bubble by a disruption of DNA structure, in which the template strand of the transiently unwound region is paired with the nascent RNA at the growing point.*

- **Termination** involves recognition of the point at which no further bases should be added to the chain. To terminate transcription, the formation of phosphodiester bonds must cease, and the transcription complex must come apart. When the last base is added to the RNA chain, the transcription bubble collapses as the RNA-DNA hybrid is disrupted, the DNA reforms in duplex state, and the enzyme and RNA are both released. *The sequence of DNA required for these reactions defines the terminator.*

The traditional view of elongation has been that it is a monotonic process, in which the enzyme moves forward 1 bp along DNA for every nucleotide added to the RNA chain. Changes in this pattern occur in certain circumstances, in particular when RNA polymerase pauses. One type of pattern is for the “front end” of the enzyme to remain stationary while the “back end” continues to move, thus compressing the footprint on DNA. After movement of several base pairs, the “front end” is released, restoring a footprint of full length. This gave rise to the “inchworm” model of transcription, in which the enzyme proceeds discontinuously, alternatively compressing and releasing the footprint on DNA. It may, however, be the case that these events describe an aberrant situation rather than normal transcription.

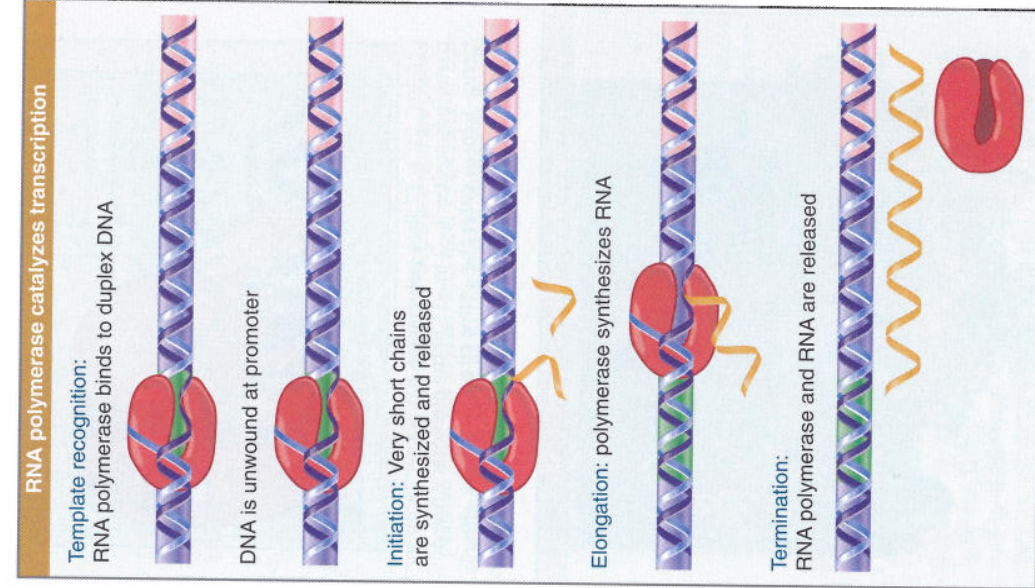


FIGURE 11.6 Transcription has four stages: The enzyme binds to the promoter and melts DNA, remains stationary during initiation, moves along the template during elongation, and dissociates at termination.

11.4 Phage T7 RNA Polymerase Is a Useful Model System

Key concepts

- T3 and T7 phage RNA polymerases are single polypeptides with minimal activities in recognizing a small number of phage promoters.
- Crystal structures of T7 RNA polymerase with DNA identify the DNA-binding region and the active site.

The existence of very small RNA polymerases, comprising single polypeptide chains coded by certain phages, gives some idea of the “mini-mum” apparatus necessary for transcription. These RNA polymerases recognize just a few promoters on the phage DNA, and they have no ability to change the set of promoters to which

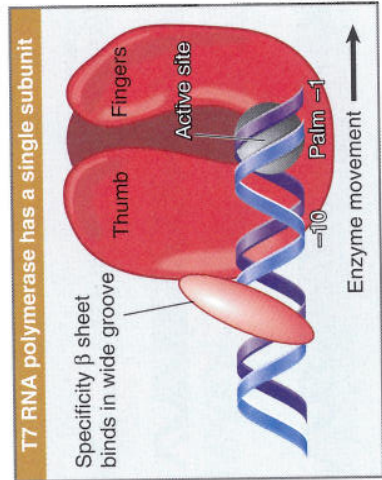


FIGURE 11.7 T7 RNA polymerase has a specificity loop that binds positions -7 to -11 of the promoter while positions -1 to -4 enter the active site.

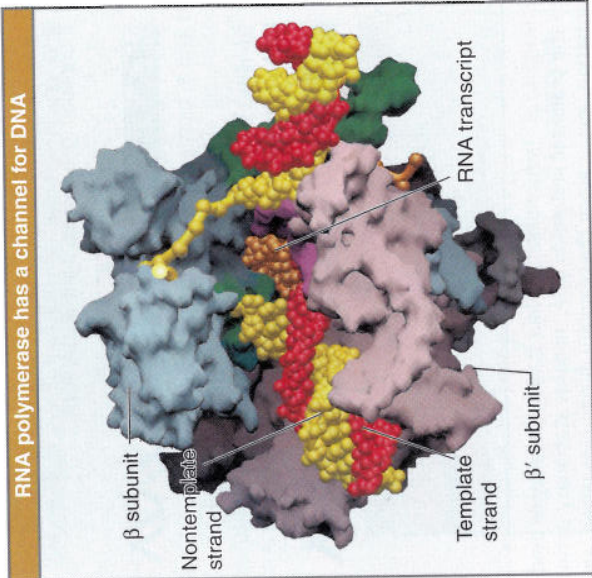


FIGURE 11.8 The β (cyan) and β' subunit (pink) of RNA polymerase have a channel for the DNA template. Synthesis of an RNA transcript (copper) has just begun; the DNA template (red) and coding (yellow) strands are separated in a transcription bubble. Photo courtesy of Seth Darst, Rockefeller University.

they respond. They provide simple model systems for characterizing the binding of RNA polymerase to DNA and the initiation reaction.

The RNA polymerases coded by the related phages T3 and T7 are single polypeptide chains of <100 kD each. They synthesize RNA at rates of ~200 nucleotides/second at 37°C, a rate that is more rapid than that of bacterial RNA polymerase.

The T7 RNA polymerase is homologous to DNA polymerases and has a similar structure, in which DNA lies in a "palm" surrounded by "fingers" and a "thumb" (see Figure 18.7). We now have a direct view of the active site from

a crystal structure of a phage T7 RNA polymerase engaged in transcription.

The T7 RNA polymerase recognizes its target sequence in DNA by binding to bases in the major groove at a position upstream from the startpoint, as shown in **FIGURE 11.7**. The enzyme uses a *specificity loop* that is formed by a β ribbon. This feature is unique to the RNA polymerase (it is not found in DNA polymerases). The common point with all RNA polymerases is that the enzyme recognizes specific bases in DNA that are upstream of the sequence that is transcribed.

When transcription initiates, the conformation of the enzyme remains essentially the same while several nucleotides are added, and the transcribed template strand is "scrunched" in the active site. The active site can hold a transcript of six to nine nucleotides. The transition from initiation to elongation is defined as the point when the enzyme begins to move along DNA. This occurs when the nascent transcript extends beyond the active site and interacts with the specificity loop. The RNA emerges to the surface of the enzyme when twelve to fourteen nucleotides have been synthesized. These features are similar to those displayed by bacterial RNA polymerase.

11.5 A Model for Enzyme Movement Is Suggested by the Crystal Structure

Key concepts

- DNA moves through a groove in yeast RNA polymerase that makes a sharp turn at the active site.
- A protein bridge changes conformation to control the entry of nucleotides to the active site.

We now have much information about the structure and function of RNA polymerase as the result of the crystal structures of the bacterial and yeast enzymes. Bacterial RNA polymerase has overall dimensions of $\sim 90 \times 95 \times 160$ Å; eukaryotic RNA polymerase is larger but less elongated. Structural analysis shows that they share a common type of structure, in which there is a "channel" or groove on the surface ~ 25 Å wide that could be the path for DNA. An example of this channel in bacterial RNA polymerase is illustrated in **FIGURE 11.8**. The length of the groove could hold 16 bp in the bacterial enzyme, and ~ 25 bp in the eukaryotic enzyme, but this represents only part of the total length of DNA bound during transcription. The enzyme sur-

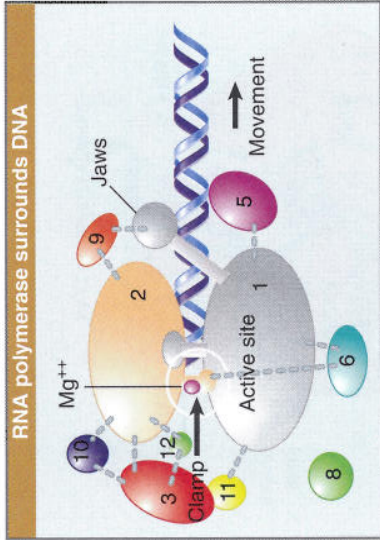


FIGURE 11.9 Ten subunits of RNA polymerase are placed in position from the crystal structure. The colors of the subunits are the same as in the crystal structures of the following figures.

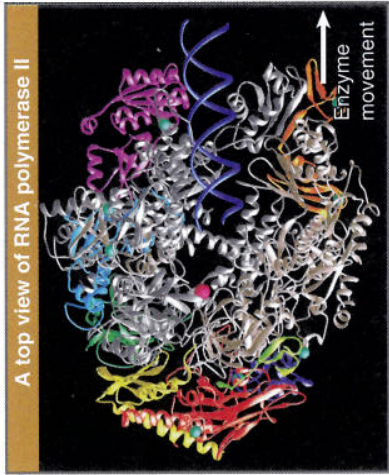


FIGURE 11.10 The top view of the crystal structure of RNA polymerase II from yeast shows that DNA is held downstream by a pair of jaws and is clamped in position in the active site, which contains an Mg^{2+} ion. Photo courtesy of Roger Kornberg, Stanford University School of Medicine.

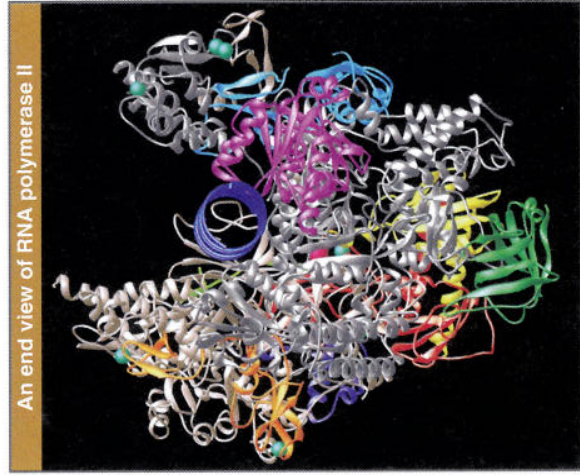


FIGURE 11.11 The end view of the crystal structure of RNA polymerase II from yeast shows that DNA is surrounded by $\sim 270^\circ$ of protein. Photo courtesy of Roger Kornberg, Stanford University School of Medicine.

face is largely negatively charged, but the groove is lined with positive charges, enabling it to interact with the negatively charged phosphate groups of DNA.

The yeast enzyme is a large structure with twelve subunits (see Section 24.2, Eukaryotic RNA Polymerases Consist of Many Subunits). Ten subunits of the yeast RNA polymerase II have been located on the crystal structure, as shown in **FIGURE 11.9**. The catalytic site is formed by a cleft between the two large subunits (#1 and #2), which grasp DNA downstream in "jaws" as it enters the RNA polymerase. Subunits 4 and 7 are missing from this structure; they form a subcomplex that dissociates from the complete enzyme. The structure is generally similar to that of bacterial RNA polymerase. This can be seen more clearly in the crystal structure of **FIGURE 11.10**. RNA polymerase surrounds the DNA, as seen in the view of **FIGURE 11.11**. A catalytic Mg^{2+} ion is found at the active site. The DNA is clamped in position at the active site by subunits 1, 2, and 6. **FIGURE 11.12** shows that DNA is forced to take a turn at the entrance to the site because of an adjacent wall of protein. The length of the RNA hybrid is limited by another protein obstruction, called the rudder. Nucleotides probably enter the active site from below, via pores through the structure.

The expanded view of the active site in **FIGURE 11.13** shows that the transcription bubble includes 9 bp of DNA-RNA hybrid. Where the DNA takes its turn, the bases downstream are flipped out of the DNA helix. As the enzyme moves along DNA, the base in the template strand at the start of the turn will be flipped to face the nucleotide entry site. The 5' end of the RNA is forced to leave the DNA when it hits the protein rudder (see Figure 11.12).

Once DNA has been melted, the individual strands have a flexible structure in the transcription bubble. This enables DNA to take its turn in the active site. Before transcription starts, though, the DNA double helix is a relatively rigid straight structure. How does this structure enter the polymerase without being blocked by the wall? The answer is that a large conformational shift must occur in the enzyme. Adjacent to the wall is a clamp. In the free form of RNA polymerase, this clamp swings away from the wall to allow DNA to follow a straight path through the enzyme. After DNA has been melted to create the transcription bubble, the clamp must swing back into position against the wall.

One of the dilemmas of any nucleic acid polymerase is that the enzyme must make tight

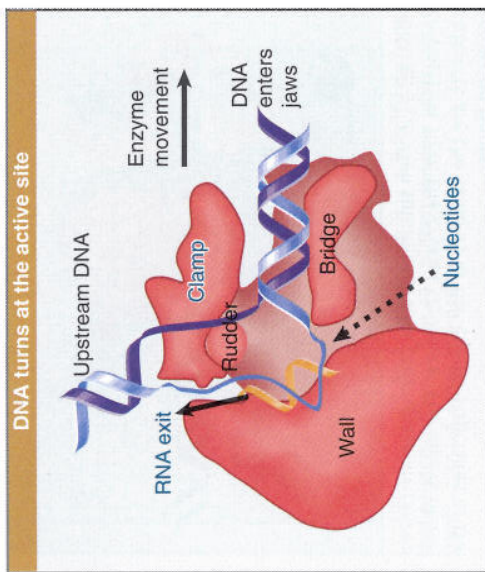


FIGURE 11.12 DNA is forced to make a turn at the active site by a wall of protein. Nucleotides may enter the active site through a pore in the protein.

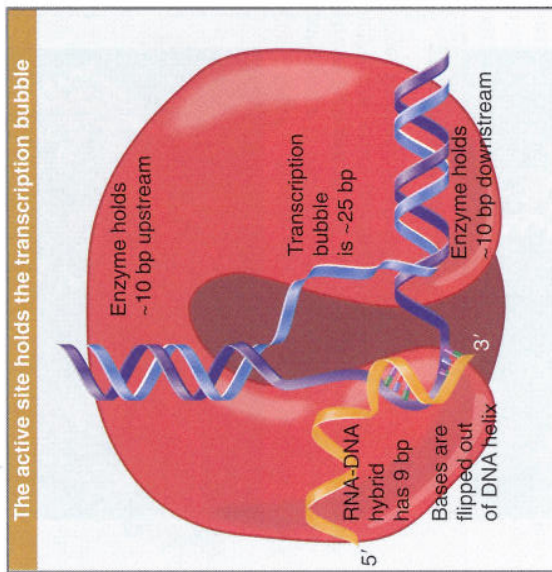


FIGURE 11.13 An expanded view of the active site shows the sharp turn in the path of DNA.

contacts with the nucleic acid substrate and product, but must break these contacts and remake them with each cycle of nucleotide addition. Consider the situation illustrated in **FIGURE 11.14**. A polymerase makes a series of specific contacts with the bases at particular positions. For example, contact "1" is made with the base at the end of the growing chain and the contact "2" is made with the base in the template strand that is complementary to the next base to be added. Note, however, that the bases that occupy these locations in the nucleic acid chains change every time a nucleotide is added!

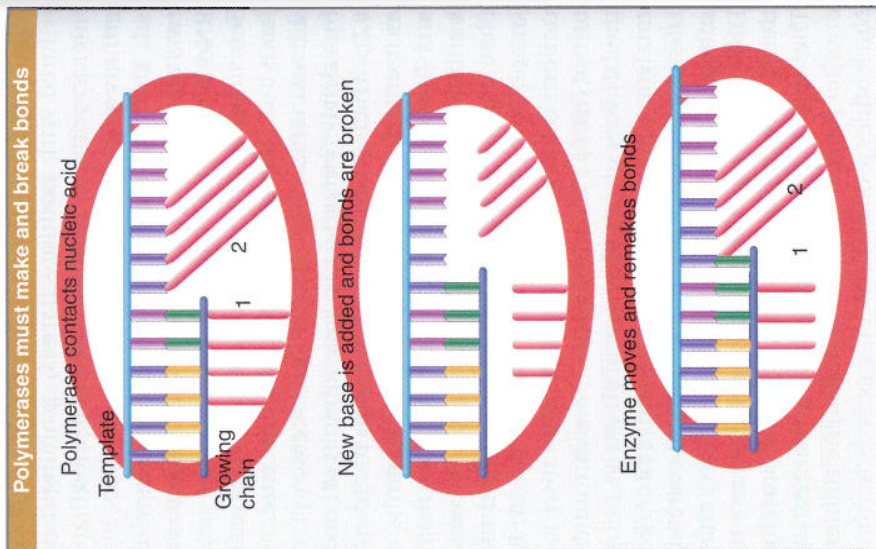


FIGURE 11.14 Movement of a nucleic acid polymerase requires breaking and remaking bonds to the nucleotides at fixed positions relative to the enzyme structure. The nucleotides in these positions change each time the enzyme moves a base along the template.

The top and bottom panels of the figure show the same situation: a base is about to be added to the growing chain. The difference is that the growing chain has been extended by one base in the bottom panel. The geometry of both complexes is exactly the same, but contacts "1" and "2" in the bottom panel are made to bases in the nucleic acid chains that are located one position farther along the chain. The middle panel shows that this must mean that, after the base is added, and before the enzyme moves relative to the nucleic acid, the contacts made to specific positions must be broken so that they can be remade to bases that occupy those positions after the movement.

The RNA polymerase structure suggests an insight into how the enzyme retains contact with its substrate while breaking and remaking bonds. A structure in the protein called the bridge is adjacent to the active site (see Figure 11.12). This feature is found in both the bacterial and yeast enzymes, but it has differ-

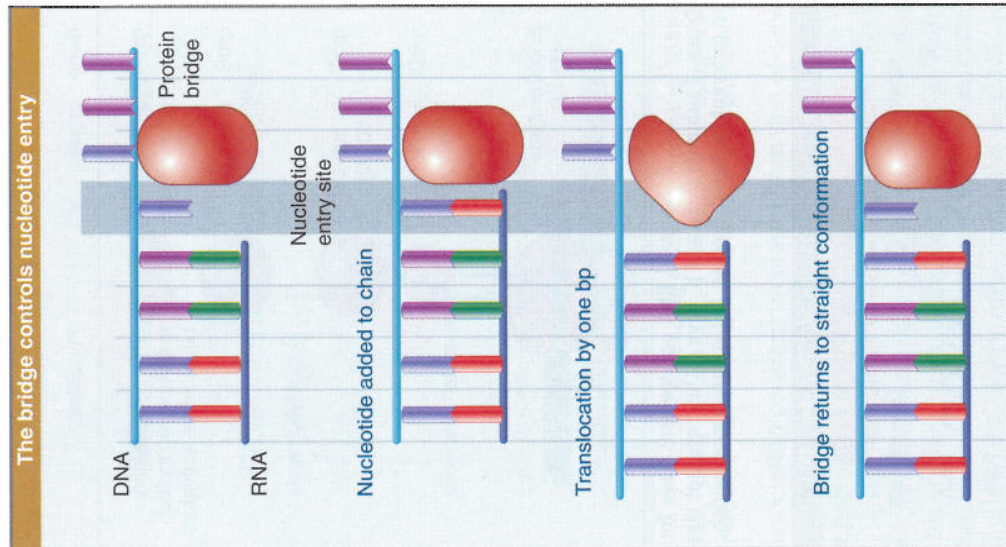


FIGURE 11.15 The RNA polymerase elongation cycle starts with a straight bridge adjacent to the nucleotide entry site. After nucleotide addition, the enzyme moves one base pair and the bridge bends as it retains contact with the newly added nucleotide. When the bridge is released, the cycle can start again.

ent shapes in the different crystal structures. In one it is bent, and in the other it is straight. **FIGURE 11.15** suggests that the change in conformation of the bridge structure is closely related to translocation of the enzyme along the nucleic acid.

At the start of the cycle of translocation, the bridge has a straight conformation adjacent to the nucleotide entry site. This allows the next nucleotide to bind at the nucleotide entry site. The bridge is in contact with the newly added nucleotide. The protein then moves one base pair along the substrate. The bridge changes its conformation, bending to keep contact with the newly added nucleotide. In this conformation, the bridge obscures the nucleotide entry site. To end the cycle, the bridge returns to its straight conformation, allowing access again to the nucleotide entry site. The bridge acts as a ratchet that releases the DNA and RNA strands for translocation while holding on to the end of the growing chain.

11.6 Bacterial RNA Polymerase Consists of Multiple Subunits

Key concepts

- Bacterial RNA core polymerases are ~500 kD multisubunit complexes with the general structure $\alpha_2\beta\beta'$.
- DNA is bound in a channel and is contacted by both the β and β' subunits.

The best characterized RNA polymerases are those of eubacteria, for which *Escherichia coli* is a typical case. A single type of RNA polymerase appears to be responsible for almost all synthesis of mRNA, and all rRNA and tRNA, in a eubacterium. About 7000 RNA polymerase molecules are present in an *E. coli* cell. Many of them are engaged in transcription; probably 2000 to 5000 enzymes are synthesizing RNA at any one time, with the number depending on the growth conditions.

The **complete enzyme** or **holoenzyme** in *E. coli* has a molecular weight of ~465 kD. Its subunit composition is summarized in **FIGURE 11.16**.

The β and β' subunits together make up the catalytic center. Their sequences are related to those of the largest subunits of eukaryotic RNA polymerases (see Section 24.2, Eukaryotic RNA Polymerases Consist of Many Subunits), suggesting that there are common features to the

actions of all RNA polymerases. The β subunit can be crosslinked to the template DNA, the product RNA, and the substrate ribonucleotides; mutations in *rpoB* affect all stages of transcription. Mutations in *rpoC* show that β' also is involved at all stages.

The α subunit is required for assembly of the core enzyme. When phage T4 infects *E. coli*, the α subunit is modified by ADP-ribosylation of an arginine. The modification is associated with a reduced affinity for the promoters formerly recognized by the holoenzyme, suggesting that the α subunit plays a role in promoter recognition. The α subunit also plays a role in the interaction of RNA polymerase with some regulatory factors.

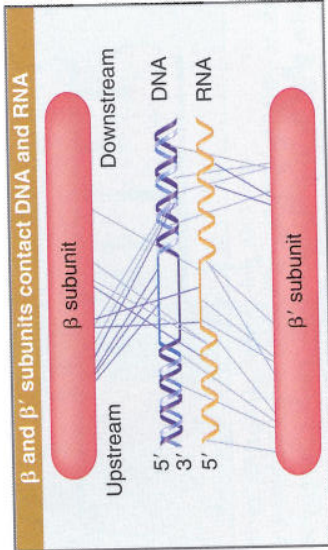
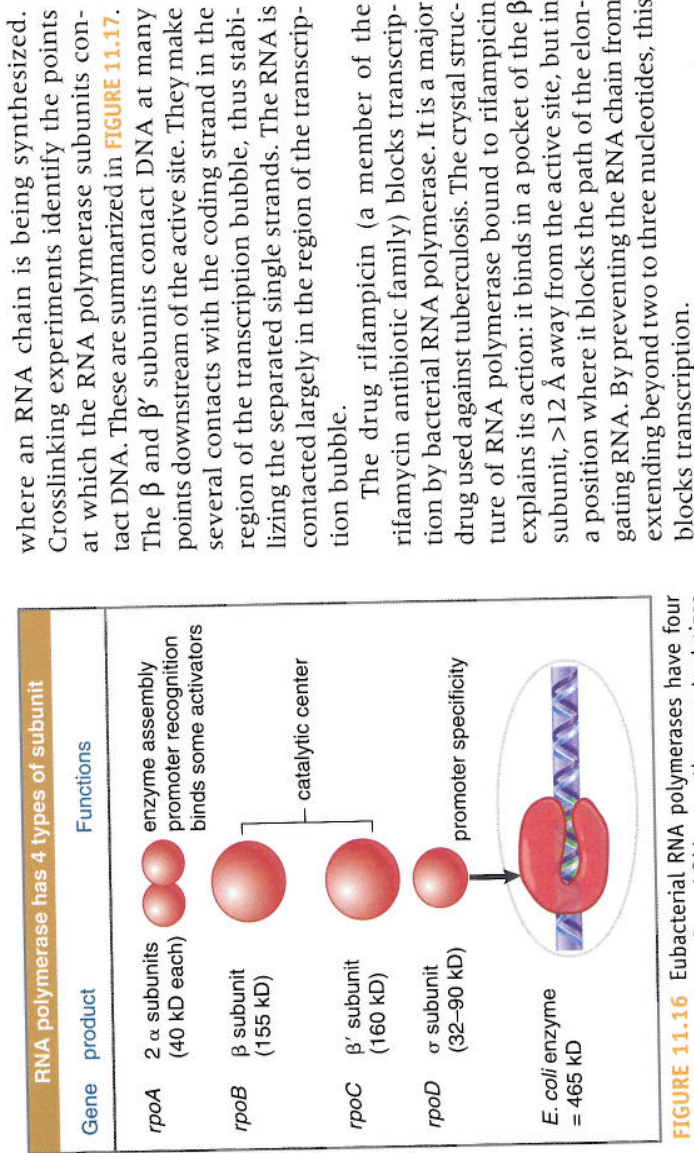


FIGURE 11.17 Both the template and coding strands of DNA are contacted by the β and β' subunits largely in the region of the transcription bubble and downstream. The RNA is contacted mostly in the transcription bubble. There is no downstream RNA, except in the special case when the enzyme backtracks.

The σ subunit is concerned specifically with promoter recognition, and we have more information about its functions than on any other subunit (see Section 11.7, RNA Polymerase Consists of the Core Enzyme and Sigma Factor).

The crystal structure of the bacterial enzyme (see Figure 11.8) shows that the channel for DNA lies at the interface of the β and β' subunits. (The α subunits are not visible in this view.) The DNA is unwound at the active site,

11.7 RNA Polymerase Consists of the Core Enzyme and Sigma Factor

Key concepts

- Bacterial RNA polymerase can be divided into the $\alpha_2\beta\beta'$ core enzyme that catalyzes transcription and the sigma subunit that is required only for initiation.
- Sigma factor changes the DNA-binding properties of RNA polymerase so that its affinity for general DNA is reduced and its affinity for promoters is increased.
- Binding constants of RNA polymerase for different promoters vary over six orders of magnitude, corresponding to the frequency with which transcription is initiated at each promoter.

The holoenzyme ($\alpha_2\beta\beta'\sigma$) can be separated into two components, the **core enzyme** ($\alpha_2\beta\beta'$) and the **sigma factor** (the σ polypeptide). Only the holoenzyme can initiate transcription. Sigma factor ensures that bacterial RNA polymerase binds in a stable manner to DNA only at promoters. The sigma “factor” is usually released when the RNA chain reaches eight to nine bases, leaving the core enzyme to undertake elongation. Core enzyme has the ability to synthesize RNA on a DNA template, but cannot initiate transcription at the proper sites.

The core enzyme has a general affinity for DNA, in which electrostatic attraction between the basic protein and the acidic nucleic acid plays a major role. Any (random) sequence of DNA that is bound by core polymerase in this general binding reaction is described as a **loose binding site**. No change occurs in the DNA, which remains duplex. The complex at such a site is stable, with a half-life for dissociation of the enzyme from DNA ~60 minutes. Core enzyme does not distinguish between promoters and other sequences of DNA.

FIGURE 11.18 shows that the sigma factor introduces a major change in the affinity of RNA polymerase for DNA. The holoenzyme has a drastically reduced ability to recognize loose binding sites—that is, to bind to any general sequence of DNA. The association constant for the reaction is reduced by a factor of $\sim 10^4$, and the half-life of the complex is <1 second. So sigma factor destabilizes the general binding ability very considerably.

Note that sigma factor also confers the ability to recognize specific binding sites. The holoenzyme binds to promoters very tightly, with an

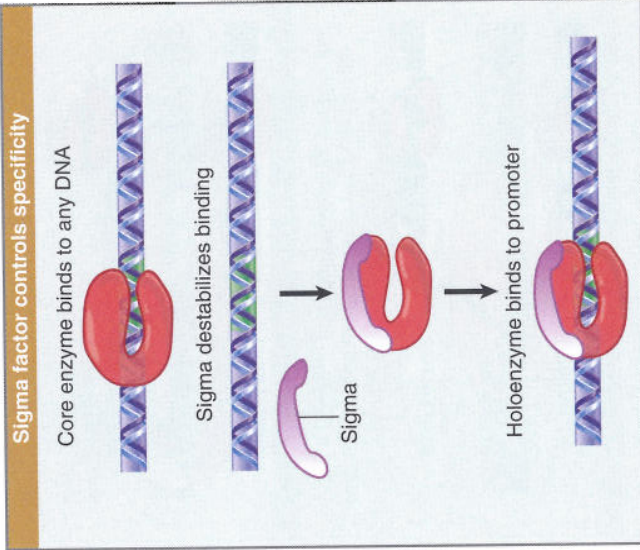


FIGURE 11.18 Core enzyme binds indiscriminately to any DNA. Sigma factor reduces the affinity for sequence-independent binding and confers specificity for promoters. association constant increased from that of core enzyme by (on average) 1000 times and with a half-life of several hours.

The specificity of holoenzyme for promoters compared to other sequences is $\sim 10^7$, but this is only an average because there is wide variation in the rate at which the holoenzyme binds to different promoter sequences. This is an important parameter in determining the efficiency of an individual promoter in initiating transcription. The binding constants range from $\sim 10^{12}$ to $\sim 10^6$. Other factors also affect the frequency of initiation, which varies from $\sim 1/\text{sec}$ (rRNA genes) to $\sim 1/30 \text{ min}$ (the *lacI* promoter).

11.8 The Association with Sigma Factor Changes at Initiation

Key concepts

- When RNA polymerase binds to a promoter, it separates the DNA strands to form a transcription bubble and incorporates up to nine nucleotides into RNA.
- There may be a cycle of abortive initiations before the enzyme moves to the next phase.
- Sigma factor may be released from RNA polymerase when the nascent RNA chain reaches eight to nine bases in length.

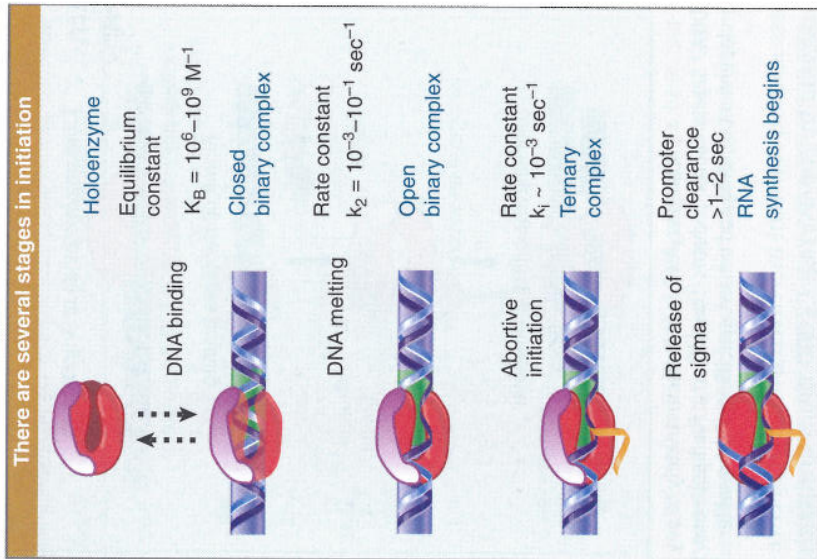


FIGURE 11.19 RNA polymerase passes through several steps prior to elongation. A closed binary complex is converted to an open form and then into a ternary complex.

We can now describe the stages of transcription in terms of the interactions between different forms of RNA polymerase and the DNA template. The initiation reaction can be described by the parameters that are summarized in **FIGURE 11.19**:

- The holoenzyme-promoter reaction starts by forming a closed binary complex. “Closed” means that the DNA remains duplex. The formation of the closed binary complex is reversible; thus it is usually described by an equilibrium constant (K_B). There is a wide range in values of the equilibrium constant for forming the closed complex.
- The closed complex is converted into an **open complex** by “melting” of a short region of DNA within the sequence bound by the enzyme. The series of events leading to formation of an open complex is called **tight binding**. For strong promoters, conversion into an open binary complex is irreversible, so this reaction is described by a rate con-

stant (k_2). This reaction is fast. Sigma factor is involved in the melting reaction (see Section 11.16, Substitution of Sigma Factors May Control Initiation). The next step is to incorporate the first two nucleotides; a phosphodiester bond then forms between them. This generates a **ternary complex** that contains RNA as well as DNA and enzyme. Formation of the ternary complex is described by the rate constant k_1 ; this is even faster than the rate constant k_2 . Further nucleotides can be added without any enzyme movement to generate an RNA chain of up to nine bases. After each base is added, there is a certain probability that the enzyme will release the chain. This comprises an **abortive initiation**, after which the enzyme begins again with the first base. A cycle of abortive initiations usually occurs to generate a series of very short oligonucleotides.

- When initiation succeeds, sigma is no longer necessary, and the enzyme makes the transition to the elongation ternary complex of core polymerase-DNA nascent RNA. The critical parameter here is *how long it takes for the polymerase to leave the promoter so another polymerase can initiate*. This parameter is the promoter clearance time; its minimum value of one to two seconds establishes the maximum frequency of initiation as <1 event per second. The enzyme then moves along the template, and the RNA chain extends beyond ten bases.
- When RNA polymerase binds to DNA, the elongated dimension of the protein extends along the DNA, but some interesting changes in shape occur during transcription. Transitions in shape and size identify three forms of the complex, as illustrated in **FIGURE 11.20**:

11.9 A Stalled RNA Polymerase Can Restart

Key concept

- An arrested RNA polymerase can restart transcription by cleaving the RNA transcript to generate a new 3' end.

RNA polymerase must be able to handle situations when transcription is blocked. This can happen, for example, when DNA is damaged. A model system for such situations is provided by arresting elongation *in vitro* by omitting one of the necessary precursor nucleotides. When the missing nucleotide is restored, the enzyme can overcome the block by cleaving the 3' end of the RNA, to create a new 3' terminus for chain elongation. The cleavage involves accessory factors in addition to the enzyme itself. In the case of *E. coli* RNA polymerase, the proteins GreA and GreB release the RNA polymerase from elongation arrest. In eukaryotic cells, RNA polymerase II requires an accessory factor ($\text{TF}_{\text{II}}\text{S}$), which enables the polymerase to cleave a few ribonucleotides from the 3' terminus of the RNA product.

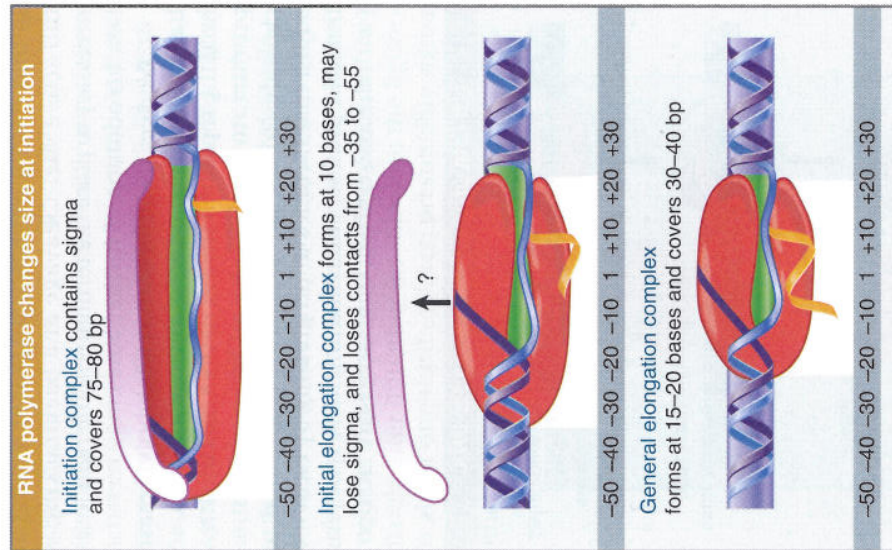


FIGURE 11.20 RNA polymerase initially contacts the region from –55 to +20. When sigma dissociates, the core enzyme contracts to –30; when the enzyme moves a few base pairs, it becomes more compactly organized into the general elongation complex.

The catalytic site of RNA polymerase undertakes the actual cleavage in each case. The roles of GreB and $\text{TF}_{\text{II}}\text{S}$ are to convert the enzyme's catalytic site into a ribonucleolytic site. Even though there is no sequence homology between the factors, crystal structures of their complexes with the respective RNA polymerases suggest that they function in a similar way. Each of the factors inserts a narrow protein domain (in one case a zinc ribbon, in the other a coiled coil) deep into RNA polymerase, where it terminates within the catalytic site. The inserted domain positions two acidic amino acids close to the primary catalytic magnesium ion of the active site; this allows the introduction of a second magnesium ion, which converts the catalytic site to a ribonucleolytic site.

The reason for this reaction may be that stalling causes the template to be mispositioned,

so that the 3' terminus is no longer located in the active site. Cleavage and backtracking is necessary to place the terminus in the right location for addition of further bases.

We see, therefore, that RNA polymerase has the facility to unwind and rewind DNA, to hold the separated strands of DNA and the RNA product, to catalyze the addition of ribonucleotides to the growing RNA chain, and to adjust to difficulties in progressing by cleaving the RNA product and restarting RNA synthesis (with the assistance of some accessory factors).

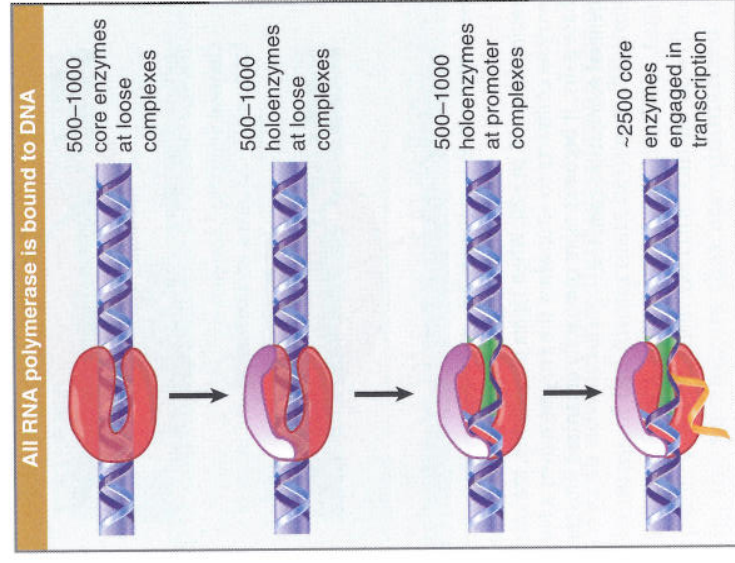


FIGURE 11.21 Core enzyme and holoenzyme are distributed on DNA, and very little RNA polymerase is free.

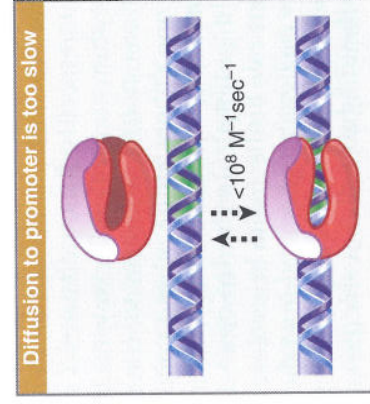


FIGURE 11.22 The forward rate constant for RNA polymerase binding to promoters is faster than random diffusion.

11.10 How Does RNA Polymerase Find Promoter Sequences?

Key concepts

- The rate at which RNA polymerase binds to promoters is too fast to be accounted for by random diffusion.
- RNA polymerase probably binds to random sites on DNA and exchanges them with other sequences very rapidly until a promoter is found.

How is RNA polymerase distributed in the cell? A (somewhat speculative) picture of the enzyme's situation is depicted in **FIGURE 11.21**:

- Excess core enzyme exists largely as closed loose complexes because the enzyme enters into them rapidly and leaves them slowly. There is very little, if any, free core enzyme.
- There is enough sigma factor for about one third of the polymerases to exist as holoenzymes, and they are distributed between loose complexes at nonspecific sites and binary complexes (mostly closed) at promoters.
- About half of the RNA polymerases consist of core enzymes engaged in transcription.
- How much holoenzyme is free? We do not know, but we suspect that the amount is very small.

RNA polymerase must find promoters within the context of the genome. Suppose that a promoter is a stretch of ~60 bp. How is it distinguished from the 4×10^6 bp that comprise the *E. coli* genome? The next three figures illustrate the principle of some possible models.

FIGURE 11.22 shows the simplest model for promoter binding, in which RNA polymerase moves by random diffusion. Holoenzyme very rapidly associates with, and dissociates from, loose binding sites. Thus it could continue to make and break a series of closed complexes until (by chance) it encounters a promoter, and its recognition of the specific sequence would allow tight binding to occur by formation of an open complex.

For RNA polymerase to move from one binding site on DNA to another, it must dissociate from the first site, find the second site, and then associate with it. Movement from one site to another is limited by the speed of diffusion through the medium. Diffusion sets an upper limit for the rate constant for associating with

a 60 bp target of $<10^8 \text{ M}^{-1} \text{ sec}^{-1}$. The actual forward rate constant for some promoters *in vitro*, however, appears to be $\sim 10^8 \text{ M}^{-1} \text{ sec}^{-1}$, at or above the diffusion limit. If this value applies *in vivo*, the time required for random cycles of successive association and dissociation at loose binding sites is too great to account for the way RNA polymerase finds its promoter.

RNA polymerase must therefore use some other means to seek its binding sites. **FIGURE 11.23** shows that the process could be speeded up if the initial target for RNA polymerase is the whole genome, not just a specific promoter sequence. By increasing the target size, the rate constant for diffusion to DNA is correspondingly increased and is no longer limiting.

If this idea is correct, a free RNA polymerase binds DNA and then remains in contact with it. How does the enzyme move from a random (loose) binding site on DNA to a promoter? The most likely model is to suppose that the bound sequence is directly displaced by another sequence. Having taken hold of DNA, the enzyme exchanges this sequence with another sequence very rapidly and continues to exchange sequences until a promoter is found. The enzyme then forms a stable, open complex, after which initiation occurs. The search process becomes much faster because association and dissociation are virtually simultaneous and time is not spent commuting between sites. Direct displacement can give a "directed walk," in which the enzyme moves preferentially from a weak site to a stronger site.

Another idea supposes that the enzyme slides along the DNA by a one-dimensional random walk, as shown in **FIGURE 11.24**, being halted only when it encounters a promoter. There is, however, no evidence that RNA polymerase (or other DNA-binding proteins) can function in this manner.

11.11 Sigma Factor Controls Binding to DNA

Key concept

- A change in association between sigma factor and holoenzyme changes binding affinity for DNA so that core enzyme can move along DNA.

RNA polymerase encounters a dilemma in reconciling its needs for initiation with those for elongation. Initiation requires tight binding *only* to particular sequences (promoters), whereas elongation requires close association with *all*

sequences that the enzyme encounters during transcription. **FIGURE 11.25** illustrates how the dilemma is solved by the reversible association between sigma factor and core enzyme. As mentioned previously (see Section 11.8, The Association with Sigma Factor Changes at Initiation), sigma factor is either released following initiation or changes its association with core enzyme so that it no longer participates in DNA binding. There are fewer molecules of sigma than of core enzyme; thus the utilization of core enzyme requires that sigma recycles. This occurs immediately after initiation (as shown in the figure) in about one third of cases; presumably sigma and core dissociate at some later point in the other cases.

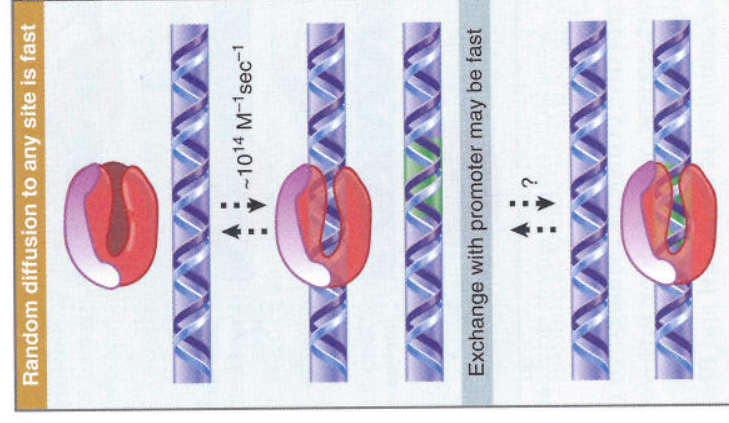


FIGURE 11.23 RNA polymerase binds very rapidly to random DNA sequences and could find a promoter by direct displacement of the bound DNA sequence.

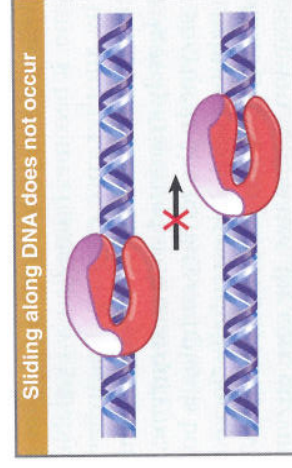


FIGURE 11.24 RNA polymerase does not slide along DNA.

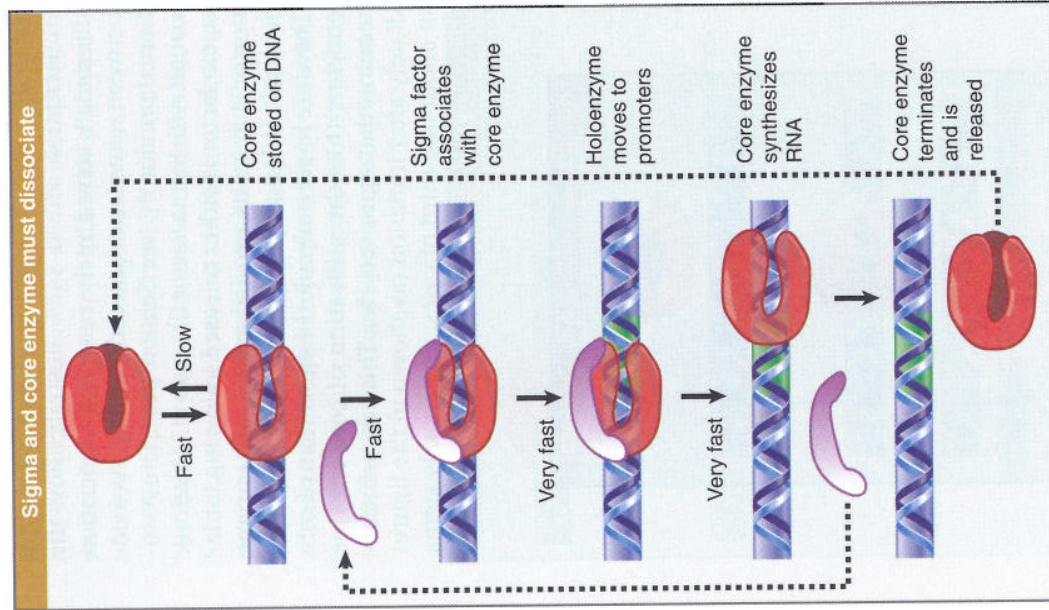


FIGURE 11.25 Sigma factor and core enzyme recycle at different points in transcription.

Irrespective of the exact timing of its release from core enzyme, sigma factor is involved only in initiation. It becomes unnecessary when abortive initiation is concluded and RNA synthesis has been successfully initiated. We do not know whether the state of polymerase changes as a consequence of overcoming abortive initiation, or whether instead it is the change in state that ends abortive initiation and allows elongation to commence.

When sigma factor is released from core enzyme, it becomes immediately available for use by another core enzyme. Regardless of whether sigma factor is released or remains more loosely associated with core enzyme, the core enzyme in the ternary complex is bound very tightly to DNA. It is essentially "locked in" until elongation has been completed. When transcription terminates, the core enzyme is released. It is then "stored" by binding to a loose

site on DNA. If it has lost its sigma factor, it must find another sigma factor in order to undertake a further cycle of transcription.

Core enzyme has a high intrinsic affinity for DNA, which is increased by the presence of nascent RNA. Its affinity for loose binding sites is, however, too high to allow the enzyme to distinguish promoters efficiently from other sequences. By reducing the stability of the loose complexes, sigma factor allows the process to occur much more rapidly; and by stabilizing the association at tight binding sites, the factor drives the reaction irreversibly into the formation of open complexes. When the enzyme releases sigma factor (or changes its association with it), it reverts to a general affinity for all DNA, irrespective of sequence, that suits it to continue transcription.

What is responsible for the ability of holoenzyme to bind specifically to promoters? Sigma factor has domains that recognize the promoter DNA. As an independent polypeptide, sigma factor does not bind to DNA, but when holoenzyme forms a tight binding complex, σ contacts the DNA in the region upstream of the start point. This difference is due to a change in the conformation of sigma factor when it binds to core enzyme. The N-terminal region of free sigma factor suppresses the activity of the DNA-binding region; when sigma binds to core, this inhibition is released, and it becomes able to bind specifically to promoter sequences (see also Figure 11.35 in Section 11.17). The inability of free sigma factor to recognize promoter sequences may be important: if σ could freely bind to promoters, it might block holoenzyme from initiating transcription.

11.12 Promoter Recognition Depends on Consensus Sequences

Key concepts

- A promoter is defined by the presence of short consensus sequences at specific locations.
- The promoter consensus sequences consist of a purine at the startpoint, the hexamer TATAAT centered at -10 , and another hexamer centered at -35 .
- Individual promoters usually differ from the consensus at one or more positions.

As a sequence of DNA whose function is to be recognized by proteins, a promoter differs from sequences whose role is to be transcribed or

translated. The information for promoter function is provided directly by the DNA sequence: its structure is the signal. This is a classic example of a *cis*-acting site, as defined previously in Figure 2.16 and Figure 2.17. By contrast, expressed regions gain their meaning only after the information is transferred into the form of some other nucleic acid or protein.

A key question in examining the interaction between an RNA polymerase and its promoter is how the protein recognizes a specific promoter sequence. Does the enzyme have an active site that distinguishes the chemical structure of a particular sequence of bases in the DNA double helix? How specific are its requirements?

One way to design a promoter would be for a particular sequence of DNA to be recognized by RNA polymerase. Every promoter would consist of, or at least include, this sequence. In the bacterial genome, the minimum length that could provide an adequate signal is 12 bp. (Any shorter sequence is likely to occur—just by chance—a sufficient number of additional times to provide false signals. The minimum length required for unique recognition increases with the size of genome.) The 12 bp sequence need not be contiguous. If a specific number of base pairs separates two constant shorter sequences, their combined length could be less than 12 bp, because the *distance* of separation itself provides a part of the signal (even if the intermediate sequence is itself irrelevant).

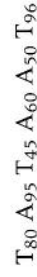
Attempts to identify the features in DNA that are necessary for RNA polymerase binding started by comparing the sequences of different promoters. Any essential nucleotide sequence should be present in all the promoters. Such a sequence is said to be **conserved**. However, a conserved sequence need not necessarily be conserved at every single position; some variation is permitted. How do we analyze a sequence of DNA to determine whether it is sufficiently conserved to constitute a recognizable signal?

Putative DNA recognition sites can be defined in terms of an idealized sequence that represents the base most often present at each position. A **consensus sequence** is defined by aligning all known examples so as to maximize their homology. For a sequence to be accepted as a consensus, each particular base must be reasonably predominant at its position, and most of the actual examples must be related to the consensus by only one or two substitutions. The striking feature in the sequence of promoters in *E. coli* is the *lack of any extensive con-*

variation of sequence over the 60 bp associated with RNA polymerase. The sequence of much of the binding site is irrelevant. Some short stretches within the promoter are conserved, however, and they are critical for its function. *Conservation of only very short consensus sequences is a typical feature of regulatory sites (such as promoters) in both prokaryotic and eukaryotic genomes.*

There are four (perhaps five) conserved features in a bacterial promoter: the startpoint, the -10 sequence, the -35 sequence, the separation between the -10 and -35 sequences, and (sometimes) the UP element:

- The startpoint is usually (>90% of the time) a purine. It is common for the startpoint to be the central base in the sequence CAT, but the conservation of this triplet is not great enough to regard it as an obligatory signal.
- Just upstream of the startpoint, a 6-bp region is recognizable in almost all promoters. The center of the hexamer generally is close to 10 bp upstream of the startpoint; the distance varies in known promoters from position -18 to -9 . Named for its location, the hexamer is often called the **-10 sequence**. Its consensus is TATAAT and can be summarized in the form



where the subscript denotes the percent occurrence of the most frequently found base, which varies from 45% to 96%. (A position at which there is no discernible preference for any base would be indicated by N.) If the frequency of occurrence indicates likely importance in binding RNA polymerase, we would expect the initial highly conserved TATA and the final almost completely conserved T in the -10 sequence to be the most important bases.

- Another conserved hexamer is centered ~ 35 bp upstream of the startpoint. This is called the **-35 sequence**. The consensus is TTGACA; in more detailed form, the conservation is



- The distance separating the -35 and -10 sites is between 16 and 18 bp in 90% of promoters; in the exceptions, it is as little as 15 bp or as great as 20 bp. *Although the actual sequence in the intervening region is unimportant, the distance is critical in*

The promoter has three components

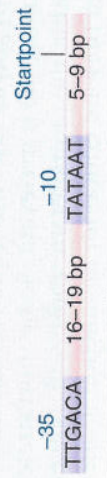


FIGURE 11.26 A typical promoter has three components, consisting of consensus sequences at -35, -10, and the startpoint.

holding the two sites at the appropriate separation for the geometry of RNA polymerase.

- Some promoters have an A-T-rich sequence located farther upstream. This is called the UP element. It interacts with the α subunit of the RNA polymerase. It is typically found in promoters that are highly expressed, such as the promoters for rRNA genes.

The optimal promoter is a sequence consisting of the -35 hexamer, separated by 17 bp from the -10 hexamer, lying 7 bp upstream of the startpoint. The structure of a promoter, showing the permitted range of variation from this optimum, is illustrated in **FIGURE 11.26**.

11.13 Promoter Efficiencies Can Be Increased or Decreased by Mutation

Key concepts

- Down mutations to decrease promoter efficiency usually decrease conformance to the consensus sequences, whereas up mutations have the opposite effect.
- Mutations in the -35 sequence usually affect initial binding of RNA polymerase.
- Mutations in the -10 sequence usually affect the melting reaction that converts a closed to an open complex.

Mutations are a major source of information about promoter function. Mutations in promoters affect the level of expression of the gene(s) they control without altering the gene products themselves. Most are identified as bacterial mutants that have lost, or have very much reduced, transcription of the adjacent genes. They are known as **down mutations**. Less often, mutants are found in which there is increased transcription from the promoter. They have **up mutations**.

It is important to remember that “up” and “down” mutations are defined relative to the usual efficiency with which a particular promoter functions. This varies widely. Thus a change that is recognized as a down mutation

in one promoter might never have been isolated in another (which in its wild-type state could be even less efficient than the mutant form of the first promoter). Information gained from studies *in vivo* simply identifies the overall direction of the change caused by mutation. Is the most effective promoter one that has the actual consensus sequences? This expectation is borne out by the simple rule that up mutations usually increase homology with one of the consensus sequences or bring the distance between them closer to 17 bp. Down mutations usually decrease the resemblance of either site with the consensus or make the distance between them more distant from 17 bp. Down mutations tend to be concentrated in the most highly conserved positions, which confirms their particular importance as the main determinant of promoter efficiency. There are, however, occasional exceptions to these rules.

To determine the absolute effects of promoter mutations, we must measure the affinity of RNA polymerase for wild-type and mutant promoters *in vitro*. There is ~100-fold variation in the rate at which RNA polymerase binds to different promoters *in vitro*, which correlates well with the frequencies of transcription when their genes are expressed *in vivo*. Taking this analysis further, we can investigate the stage at which a mutation influences the capacity of the promoter. Does it change the affinity of the promoter for binding RNA polymerase? Does it leave the enzyme able to bind but unable to initiate? Is the influence of an ancillary factor altered? By measuring the kinetic constants for formation of a closed complex and its conversion to an open complex, as defined in **Figure 11.19**, we can dissect the two stages of the initiation reaction:

- Down mutations in the -35 sequence reduce the rate of closed complex formation (they reduce K_B), but they do not inhibit the conversion to an open complex.
- Down mutations in the -10 sequence do not affect the initial formation of a closed complex, but they slow its conversion to the open form (they reduce k_2).

These results suggest the model shown in **FIGURE 11.27**. The function of the -35 sequence is to provide the signal for recognition by RNA polymerase, whereas the -10 sequence allows the complex to convert from closed to open form. We might view the -35 sequence as comprising a “recognition domain,” whereas the -10 sequence comprises an “unwinding domain” of the promoter.

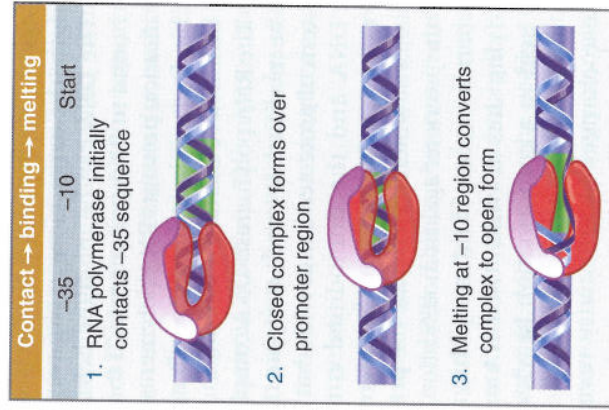


FIGURE 11.27 The -35 sequence is used for initial recognition and the -10 sequence is used for the melting reaction that converts a closed complex to an open complex.

The consensus sequence of the -10 site consists exclusively of A-T base pairs, a configuration that assists the initial melting of DNA into single strands. The lower energy needed to disrupt A-T pairs compared with G-C pairs means that a stretch of A-T pairs demands the minimum amount of energy for strand separation.

The sequence immediately around the startpoint influences the initiation event. The initial transcribed region (from +1 to +30) influences the rate at which RNA polymerase clears the promoter and therefore has an effect upon promoter strength. Thus the overall strength of a promoter cannot be predicted entirely from its -35 and -10 consensus sequences.

A “typical” promoter relies upon its -35 and -10 sequences to be recognized by RNA polymerase, but one or the other of these sequences can be absent from some (exceptional) promoters. In at least some of these cases, the promoter cannot be recognized by RNA polymerase alone; the reaction requires ancillary proteins, which overcome the deficiency in intrinsic interaction between RNA polymerase and the promoter.

11.14 RNA Polymerase Binds to One Face of DNA

Key concepts

- The consensus sequences at -35 and -10 provide most of the contact points for RNA polymerase in the promoter.
- The points of contact lie on one face of the DNA.

A protein protects a series of bonds against nuclease attack

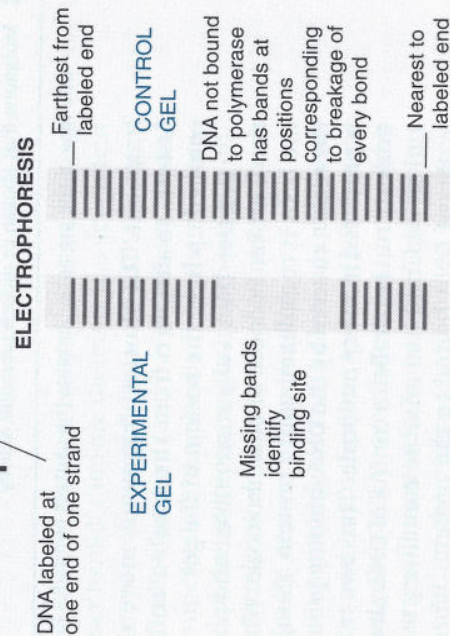
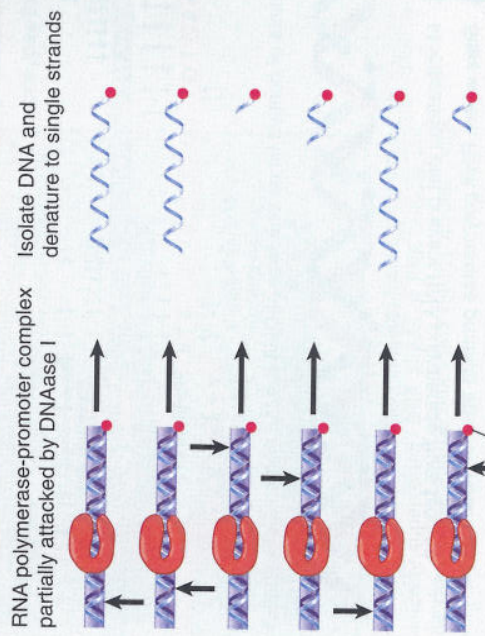


FIGURE 11.28 Footprinting identifies DNA-binding sites for proteins by their protection against nicking.

The ability of RNA polymerase (or indeed any protein) to recognize DNA can be characterized by **footprinting**. A sequence of DNA bound to the protein is *partially* digested with an endonuclease to attack individual phosphodiester bonds within the nucleic acid. Under appropriate conditions, any particular phosphodiester bond is broken in some, but not in all, DNA molecules. The positions that are cleaved are recognized by using DNA labeled on one strand at one end only. The principle is the same as that involved in DNA sequencing: partial cleavage of an end-labeled molecule at a susceptible site creates a fragment of unique length.

As **FIGURE 11.28** shows, following the nuclease treatment, the broken DNA fragments are recovered and electrophoresed on a gel that separates them according to length. Each fragment that retains a labeled end produces a radioactive band. The position of the band corresponds to the number of bases in the

with reagents that modify particular bases. We can perform the experiment in two ways:

- The DNA can be modified before it is bound to RNA polymerase. If the modification prevents RNA polymerase from binding, we have identified a base position where contact is essential.
- The RNA polymerase-DNA complex can be modified. We then compare the pattern of protected bands with that of free DNA and the unmodified complex. Some bands disappear, thus identifying sites at which the enzyme has protected the promoter against modification. Other bands increase in intensity, thus identifying sites at which the DNA must be held in a conformation in which it is more exposed.

These changes in sensitivity reveal the geometry of the complex, as summarized in **FIGURE 11.29** for a typical promoter. The regions at -35 and -10 contain most of the contact points for the enzyme. Within these regions, the same sets of positions tend both to prevent binding if previously modified, and to show increased or decreased susceptibility to modification after binding. The points of contact do not coincide completely with sites of mutation; however, they occur in the same limited region.

It is noteworthy that the same positions in different promoters provide the contact points, even though a different base is present. This indicates that there is a common mechanism for RNA polymerase binding, although the reaction does not depend on the presence of particular bases at some of the points of contact. This model explains why some of the points of contact are not sites of mutation. In addition, not every mutation lies in a point of contact; the mutations may influence the neighborhood without actually being touched by the enzyme.

It is especially significant that the experiments with prior modification identify *only* sites in the same region that is protected by the enzyme against subsequent modification. These two experiments measure different things. Prior modification identifies all those sites that the enzyme must recognize in order to bind to DNA. Protection experiments recognize all those sites that actually make contact in the binary complex. The protected sites include all the recognition sites and also some additional positions, which suggests that the enzyme first recognizes a set of bases necessary for it to "touch down" and then extends its points of contact to additional bases.

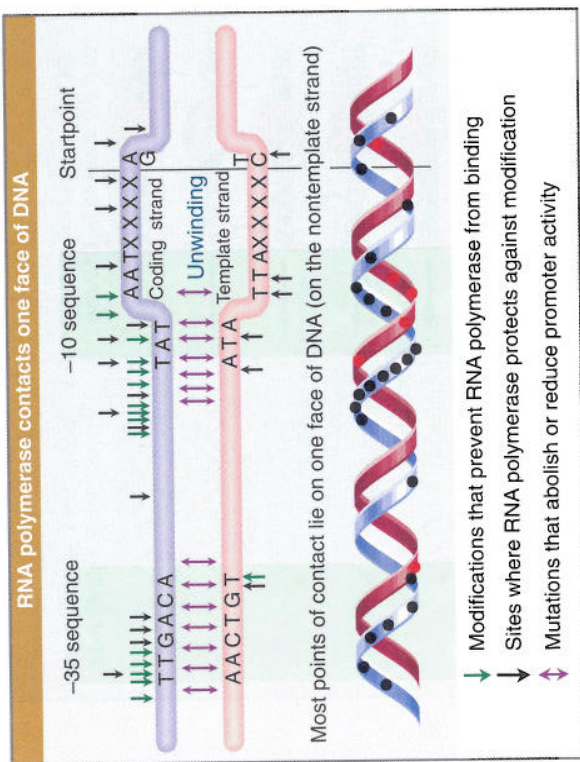


FIGURE 11.29 One face of the promoter contains the contact points for RNA.

fragment. The shortest fragments move the fastest, so distance from the labeled end is counted up from the bottom of the gel.

In a free DNA, every susceptible bond position is broken in one or another molecule. When the DNA is complexed with a protein, though, the region covered by the DNA-binding protein is protected in every molecule. Thus two reactions are run in parallel: a control of DNA alone, and an experimental mixture containing molecules of DNA bound to the protein. *When a bound protein blocks access of the nuclease to DNA, the bonds in the bound sequence fail to be broken in the experimental mixture.*

In the control, every bond is broken. This generates a series of bands, with one band representing each base. There are thirty-one bands in the figure. In the protected fragment, bonds cannot be broken in the region bound by the protein, so bands representing fragments of the corresponding sizes are not generated. The absence of bands 9–18 in the figure identifies a protein-binding site covering the region located 9–18 bases from the labeled end of the DNA. By comparing the control and experimental lanes with a sequencing reaction that is run in parallel, it becomes possible to "read off" the corresponding sequence directly, thus identifying the nucleotide sequence of the binding site.

As described previously (see Figure 11.20), RNA polymerase initially binds the region from -50 to +20. The points at which RNA polymerase actually contacts the promoter can be identified by modifying the footprinting technique to treat RNA polymerase-promoter complexes

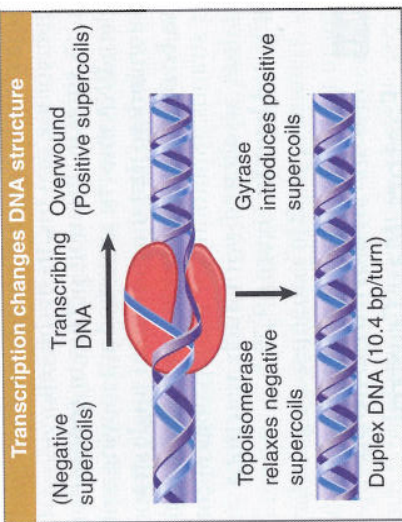


FIGURE 11.30 Transcription generates more tightly wound (positively supercoiled) DNA ahead of RNA polymerase, while the DNA behind becomes less tightly wound (negatively supercoiled).

that some promoters have sequences that are easier to melt (and are therefore less dependent on supercoiling), whereas others have more difficult sequences (and have a greater need to be supercoiled). An alternative is that the location of the promoter might be important if different regions of the bacterial chromosome have different degrees of supercoiling.

Supercoiling also has a continuing involvement with transcription. As RNA polymerase transcribes DNA, unwinding and rewinding occurs, as illustrated in Figure 11.4. This requires that either the entire transcription complex rotates about the DNA or the DNA itself must rotate about its helical axis. The consequences of the rotation of DNA are illustrated in **FIGURE 11.30** in the *twist domain* model for transcription. As RNA polymerase pushes forward along the double helix, it generates positive supercoils (more tightly wound DNA) ahead and leaves negative supercoils (partially unwound DNA) behind. For each helical turn traversed by RNA polymerase, +1 turn is generated ahead and -1 turn behind.

Transcription therefore has a significant effect on the (local) structure of DNA. As a result, the enzymes gyrase (introduces negative supercoils) and topoisomerase I (removes negative supercoils) are required to rectify the situation in front of and behind the polymerase, respectively. Blocking the activities of gyrase and topoisomerase causes major changes in the supercoiling of DNA. For example, in yeast lacking an enzyme that relaxes negative supercoils, the density of negative supercoiling doubles in a transcribed region. A possible implication of these results is that transcription is responsible

11.15 Supercoiling Is an Important Feature of Transcription

Key concepts

- Negative supercoiling increases the efficiency of some promoters by assisting the melting reaction.
- Transcription generates positive supercoils ahead of the enzyme and negative supercoils behind it, and these must be removed by gyrase and topoisomerase.

The importance of strand separation in the initiation reaction is emphasized by the effects of supercoiling. Both prokaryotic and eukaryotic RNA polymerases can initiate transcription more efficiently *in vitro* when the template is supercoiled, presumably because the supercoiled structure requires less free energy for the initial melting of DNA in the initiation complex.

The efficiency of some promoters is influenced by the degree of supercoiling. The most common relationship is for transcription to be aided by negative supercoiling. We understand in principle how this assists the initiation reaction. Why, though, should some promoters be influenced by the extent of supercoiling whereas others are not? One possibility is that the dependence of a promoter on supercoiling is determined by its sequence. This would predict

for generating a significant proportion of the supercoiling that occurs in the cell.

A similar situation occurs in replication, when DNA must be unwound at a moving replication fork so that the individual single strands can be used as templates to synthesize daughter strands. (Solutions for the topological constraints associated with such reactions are indicated later, in Figure 19.20.)

11.16 Substitution of Sigma Factors May Control Initiation

Key concepts

- E. coli* has several sigma factors, each of which causes RNA polymerase to initiate at a set of promoters defined by specific -35 and -10 sequences.
- σ^{70} is used for general transcription, and the other sigma factors are activated by special conditions.

The division of labors between a core enzyme that undertakes chain elongation and a sigma factor involved in site selection immediately raises the question of whether there is more than one type of sigma factor, each specific for a different class of promoters. **FIGURE 11.31** shows the principle of a system in which a substitution of the sigma factor changes the choice of promoter.

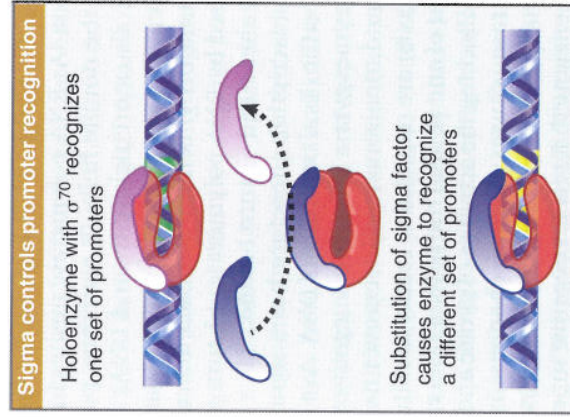


FIGURE 11.31 The sigma factor associated with core enzyme determines the set of promoters at which transcription is initiated.

E. coli uses alternative sigma factors to respond to general environmental changes; they are listed in **FIGURE 11.32**. (They are named either by molecular weight of the product or for the gene.) The general factor, which is responsible for transcription of most genes under normal conditions, is σ^{70} . The alternative sigma factors σ^S , σ^{32} , σ^E , and σ^{54} are activated in response to environmental changes; σ^{28} is used for expression of flagellar genes during normal growth, but its level of expression responds to changes in the environment. All the sigma factors except σ^{54} belong to the same protein family and function in the same general manner.

Temperature fluctuation is a common type of environmental challenge. Many organisms, both prokaryotic and eukaryotic, respond in a similar way. Upon an increase in temperature, synthesis of the proteins currently being made is turned off or down, and a new set of proteins is synthesized. The new proteins are the products of the **heat shock genes**, which play a role in protecting the cell against environmental stress. Heat shock genes are synthesized in response to conditions other than heat shock as well. Several of the heat shock proteins are chaperones. In *E. coli*, the expression of seventeen heat shock proteins is triggered by changes at transcription. The gene *rpoH* is a regulator needed to switch on the heat shock response. Its product is σ^{32} , which functions as an alternative sigma factor that causes transcription of the heat shock genes.

The heat shock response is accomplished by increasing the amount of σ^{32} when the temperature increases and decreasing its activity when the temperature change is reversed. The basic signal that induces production of σ^{32} is the accumulation of unfolded (partially denatured) proteins that results from increase in temperature. The σ^{32} protein is unstable, which is important in allowing its quantity to be increased or decreased rapidly. The proteins σ^{70} and σ^{32} can compete for the available core enzyme, so that the set of genes transcribed during heat shock depends on the balance between them. Changing sigma factors is a serious matter that has widespread implications for gene expression in the bacterium. It is not surprising, therefore, that the production of new sigma factors can be the target of many regulatory circuits. The factor σ^S is induced when bacteria make the transition from growth phase to stationary phase and also in other stress conditions. It is controlled at two levels. Translation of the *rpoS* mRNA is increased by low temperature or high

osmolarity. Proteolysis of the protein product is inhibited by carbon starvation (the typical signal of stationary phase) and by high temperature.

Another group of heat-regulated genes is controlled by the factor σ^E . It responds to more extreme temperature shifts than σ^{32} and is induced by accumulation of unfolded proteins in the periplasmic space or outer membrane. It is controlled by the intricate circuit summarized in **FIGURE 11.33**. The factor σ^E binds to a protein (RseA) that is located in the inner membrane. As a result, it cannot activate transcription. The accumulation of unfolded proteins activates a protease (DegS) in the periplasmic space, which cleaves off the C-terminal end of the RseA protein. This cleavage activates another protein in the inner membrane (YaeL), which cleaves the N-terminal region of RseA. When this happens, the σ^E factor is released and can then activate transcription. The net result is that the accumulation of unfolded proteins at the periphery of the bacterium is responsible for activating the set of genes controlled by the sigma factor.

This circuit has two interesting parallels with other regulatory circuits. The response to unfolded proteins in eukaryotic cells also uses a pathway in which an unfolded protein (within the endoplasmic reticulum) activates a membrane protein. In this case, the membrane protein is an endonuclease that cleaves an RNA, leading ultimately to a change in splicing that causes the production of a transcription factor (see Section 26.17, The Unfolded Protein Response Is Related to tRNA Splicing). A more direct parallel is with the first case to be discovered, in which cleavage of a membrane protein activates a transcription factor. In this case, the transcription factor itself is synthesized as a membrane protein, and the level of sterols in the membrane controls the activation of proteases that release the transcription factor from the cytosolic domain of the protein.

Another sigma factor is used under conditions of nitrogen starvation. *E. coli* cells contain a small amount of σ^{54} , which is activated when ammonia is absent from the medium. In these conditions, genes are turned on to allow utilization of alternative nitrogen sources. Counterparts to this sigma factor have been found in a wide range of bacteria, so it represents a response mechanism that has been conserved in evolution.

Another case of evolutionary conservation of sigma factors is presented by the factor σ^F ,

which is present in small amounts and causes RNA polymerase to transcribe genes involved in chemotaxis and flagellar structure. Its counterpart in *Bacillus subtilis* is σ^D , which controls flagellar and motility genes; factors with the same promoter specificity are present in many species of bacteria.

Each sigma factor causes RNA polymerase to initiate at a particular set of promoters. By analyzing the sequences of these promoters, we can show that each set is identified by unique sequence elements. Indeed, the sequence of each type of promoter ensures that it is recognized only by RNA polymerase directed by the appropriate sigma factor. We can deduce the general rules for promoter recognition from the identification of the genes responding to the sigma factors found in *E. coli* and those involved

<i>E. coli</i> has several sigma factors			
Gene	Factor	Use	
<i>rpoD</i>	σ^{70}	general	
<i>rpoS</i>	σ^S	stress	
<i>rpoH</i>	σ^{32}	heat shock	
<i>rpoE</i>	σ^E	heat shock	
<i>rpoN</i>	σ^{54}	nitrogen starvation	
<i>flhA</i>	σ^{28} (σ^F)	flagellar synthesis	

FIGURE 11.32 In addition to σ^{70} , *E. coli* has several sigma factors that are induced by particular environmental conditions. (A number in the name of a factor indicates its mass.)

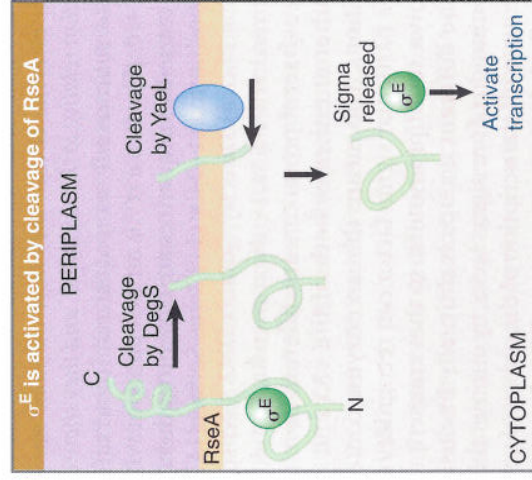


FIGURE 11.33 RseA is synthesized as a protein in the inner membrane. Its cytoplasmic domain binds the σ^E factor. RseA is cleaved sequentially in the periplasmic space and then in the cytoplasm. The cytoplasmic cleavage releases σ^E .

11.17 Sigma Factors Directly Contact DNA

Key concepts

- σ^{70} changes its structure to release its DNA-binding regions when it associates with core enzyme.
- σ^{70} binds both the -35 and -10 sequences.

Sigma factors recognize promoters by consensus sequences				
Gene	Factor	-35 Sequence	Separation	-10 Sequence
<i>rpoD</i>	σ^{70}	TTGACA	16–18 bp	TATAAT
<i>rpoH</i>	σ^{32}	CCCTTGAA	13–15 bp	CCCGATNT
<i>rpoN</i>	σ^{54}	CTGGNA	6 bp	TTGCA
<i>flaA</i>	σ^{28} (σ^F)	CTAAA	15 bp	GCCGATAA
<i>sigH</i>	σ^H	AGGANPuPu	11–12 bp	GCTGAATCA

FIGURE 11.34 *E. coli* sigma factors recognize promoters with different consensus sequences.

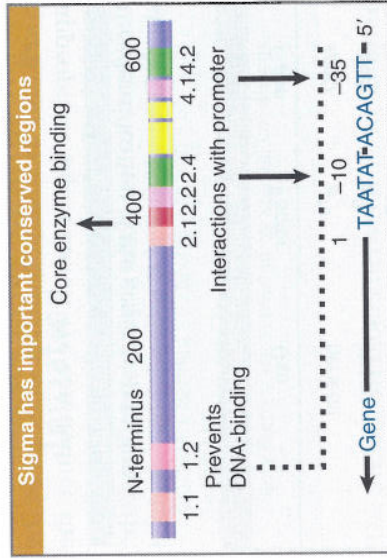


FIGURE 11.35 A map of the *E. coli* σ^{70} factor identifies conserved regions. Regions 2.1 and 2.2 contact core polymerase, 2.3 is required for melting, and 2.4 and 4.2 contact the -10 and -35 promoter elements. The N-terminal region prevents 2.4 and 4.2 from binding to DNA in the absence of core enzyme.

in sporulation in *B. subtilis* (see Section 11.19, Sporulation Is Controlled by Sigma Factors).

A significant feature of the promoters for each enzyme is that they have the same size and location relative to the startpoint, and they show conserved sequences only around the usual centers of -35 and -10. (The factor σ^{54} is an exception for which the consensus sequences are closer together and are positioned at -24 and -12; see Section 11.17, Sigma Factors Directly Contact DNA.) As summarized in **FIGURE 11.34**, the consensus sequences for each set of promoters are different from one another at either or both of the -35 and -10 positions. This means that an enzyme containing a particular sigma factor can recognize only its own set of promoters, so that transcription of the different groups is mutually exclusive. Substitution of one sigma factor by another therefore turns off transcription of the old set of genes as well as turning on transcription of a new set of genes. (Some genes are expressed by RNA polymerases with different sigma factors because they have more than one promoter, each with a different set of consensus sequences.)

(see Section 14.11, Repressor Uses a Helix-Turn-Helix Motif to Bind DNA). Amino acids separated by three to four positions lie on the same face of an α -helix and are therefore in a position to contact adjacent base pairs. **FIGURE 11.36** shows that amino acids lying along one face of the 2.4 region α -helix contact the bases at positions -12 to -10 of the -10 promoter sequence.

Region 2.3 resembles proteins that bind single-stranded nucleic acids and is involved in the melting reaction. Regions 2.1 and 2.2 (which comprise the most highly conserved part of sigma) are involved in the interaction with core enzyme. It is assumed that all sigma factors bind the same regions of the core polymerase, which ensures that the reactions are competitive.

The N-terminal region of σ^{70} has important regulatory functions. If it is removed, the shortened protein becomes able to bind specifically to promoter sequences. This suggests that the N-terminal region behaves as an autoinhibition domain. It occludes the DNA-binding domains when σ^{70} is free. Association with core enzyme changes the conformation of sigma so that the inhibition is released, and the DNA-binding domains can contact DNA.

FIGURE 11.37 schematizes the conformational change in sigma factor at open complex formation. When sigma factor binds to the core polymerase, the N-terminal domain swings ~20 Å away from the DNA-binding domains, and the DNA-binding domains separate from one another by ~15 Å, presumably to acquire a more elongated conformation appropriate for contacting DNA. Mutations in either the -10 or -35 sequences prevent an (N-terminal-deleted) σ^{70} from binding to DNA, which suggests that σ^{70} contacts both sequences simultaneously. This implies that the sigma factor must have a rather elongated structure, extending over the ~68 Å of two turns of DNA.

In the free holoenzyme, the N-terminal domain is located in the active site of the core enzyme components, essentially mimicking the location that DNA will occupy when a transcription complex is formed. When the holoenzyme forms an open complex on DNA, the N-terminal sigma domain is displaced from the active site. Its relationship with the rest of the protein is therefore very flexible; the relationship changes when sigma factor binds to core enzyme and again when the holoenzyme binds to DNA.

Comparisons of the crystal structures of the core enzyme and holoenzyme show that

The 2.4 helix of sigma determines specificity

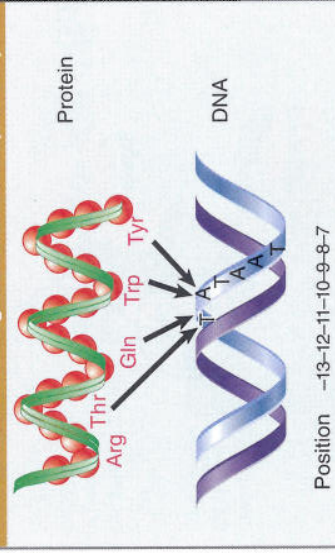


FIGURE 11.36 Amino acids in the 2.4 α -helix of σ^{70} contact specific bases in the coding strand of the -10 promoter sequence.

Sigma N-terminus controls DNA-binding

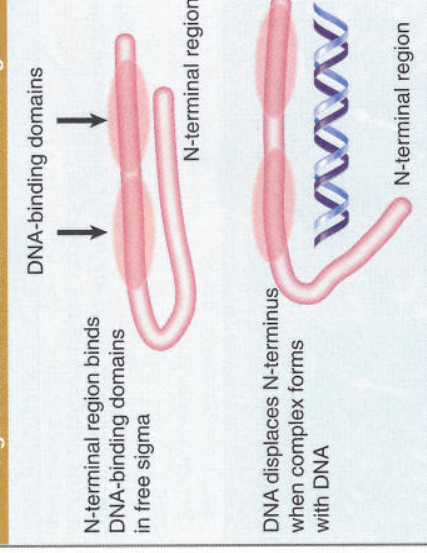


FIGURE 11.37 The N-terminus of sigma blocks the DNA-binding regions from binding to DNA. When an open complex forms, the N-terminus swings 20 Å away, and the two DNA-binding regions separate by 15 Å.

sigma factor lies largely on the surface of the core enzyme. **FIGURE 11.38** shows that it has an elongated structure that extends past the DNA-binding site. This places it in a position to contact DNA during the initial binding. The DNA helix has to move some 16 Å from the initial position in order to enter the active site. **FIGURE 11.39** illustrates this movement, looking in cross-section down the helical axis of the DNA.

An interesting difference in behavior is found with the σ^{54} factor. This causes RNA polymerase to recognize promoters that have a distinct consensus sequence, with a conserved element at -12 and another close by at -24 (given in the "-35" column of **Figure 11.32**). Thus the geometry of the polymerase-promoter complex is different under the direction of this sigma factor. Another difference in the

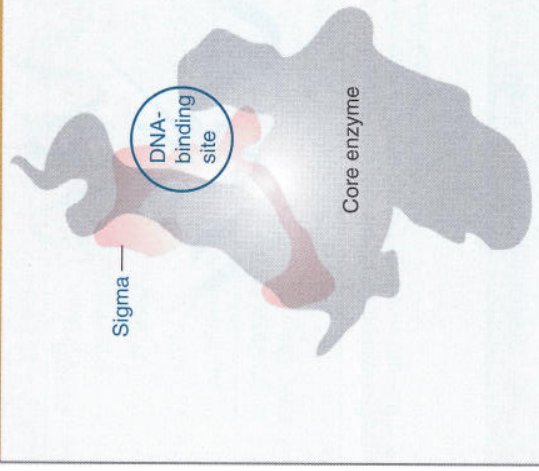


FIGURE 11.38 Sigma factor has an elongated structure that extends along the surface of the core subunits when the holoenzyme is formed.



FIGURE 11.39 DNA initially contacts sigma factor (pink) and core enzyme (gray). It moves deeper into the core enzyme to make contacts at the -10 sequence. When sigma is released, the width of the passage containing DNA increases. Reproduced from Vassilyev, D. G., et al. *Nature*. 2002. 417: 712–719. Photo courtesy of Shigeyuki Yokoyama, The University of Tokyo.

mechanism of regulation is that high-level transcription directed by σ^{54} requires other activators to bind to sites that are quite distant from the promoter. This contrasts with the other types of bacterial promoter, for which the regulator sites are always in close proximity to the promoter. The behavior of σ^{54} itself is different from other sigma factors, most notably in its ability to bind to DNA independently of core polymerase. In this regard, σ^{54} is more like the eukaryotic regulators we discuss in Chapter 24, Promoters and Enhancers, than the typical prokaryotic regulators discussed in Chapter 12, The Operon.

11.18 Sigma Factors May Be Organized into Cascades

Key concepts

- A cascade of sigma factors is created when one sigma factor is required to transcribe the gene coding for the next sigma factor.
- The early genes of phage SP01 are transcribed by host RNA polymerase.
- One of the early genes codes for a sigma factor that causes RNA polymerase to transcribe the middle genes.
- Two of the middle genes code for subunits of a sigma factor that causes RNA polymerase to transcribe the late genes.

Sigma factors are used extensively to control initiation of transcription in the bacterium *B. subtilis*, for which ~10 different σ factors are known. Some are present in vegetative cells; others are produced only in the special circumstances of phage infection or the change from vegetative growth to sporulation.

The major RNA polymerase found in *B. subtilis* cells engaged in normal vegetative growth has the same structure as that of *E. coli*, σ^{70} . Its sigma factor (described as σ^{43} or σ^A) recognizes promoters with the same consensus sequences used by the *E. coli* enzyme under direction from σ^{70} . Several variants of the RNA polymerase that contain other sigma factors are found in much smaller amounts. The variant enzymes recognize different promoters on the basis of consensus sequences at -35 and -10 .

Transitions from expression of one set of genes to expression of another set are a common feature of bacteriophage infection. In all but the very simplest cases, the development of the phage involves shifts in the pattern of transcription during the infective cycle. These shifts may be accomplished by the synthesis of a phage-encoded RNA polymerase or by the efforts of phage-encoded ancillary factors that control the bacterial RNA polymerase. A well-characterized example of control via the production of new sigma factors occurs during infection of *B. subtilis* by phage SP01.

The infective cycle of SP01 passes through three stages of gene expression. Immediately on infection, the **early genes** of the phage are transcribed. After four to five minutes, the early genes cease transcription and the **middle genes** are transcribed. At eight to twelve minutes, middle gene transcription is replaced by transcription of **late genes**.

The early genes are transcribed by the holoenzyme of the host bacterium. They are

essentially indistinguishable from host genes whose promoters have the intrinsic ability to be recognized by the RNA polymerase σ^{43} .

Expression of phage genes is required for the transitions to middle and late gene transcription. Three regulatory genes, 28, 33, and 34, control the course of transcription. Their functions are summarized in **FIGURE 11.40**. The pattern of regulation creates a **cascade**, in which the host enzyme transcribes an early gene whose product is needed to transcribe the middle genes. After this transcription, two of the middle genes code for products that are needed to transcribe the late genes.

Mutants in the early gene 28 cannot transcribe the middle genes. The product of gene 28 (called gp28) is a protein of 26 kD that replaces the host sigma factor on the core enzyme. *This substitution is the sole event required to make the transition from early to middle gene expression*. It creates a holoenzyme that can no longer transcribe the host genes but instead specifically transcribes the middle genes. We do not know how gp28 displaces σ^{43} or what happens to the host sigma polypeptide.

Two of the middle genes are involved in the next transition. Mutations in either gene 33 or 34 prevent transcription of the late genes. The products of these genes form a dimer that replaces gp28 on the core polymerase. Again, we do not know how gp33 and gp34 exclude gp28 (or any residual host σ^{43}), but *once they have bound to the core enzyme, it is able to initiate transcription only at the promoters for late genes*.

The successive replacements of sigma factors have dual consequences. Each time the subunit is changed, the RNA polymerase becomes able to recognize a new class of genes and it no longer recognizes the previous class. These switches therefore constitute global changes in the activity of RNA polymerase. Probably all, or virtually all, of the core enzyme becomes associated with the sigma factor of the moment, and the change is irreversible.

11.19 Sporulation Is Controlled by Sigma Factors

Key concepts

- Sporulation divides a bacterium into a mother cell that is lysed and a spore that is released.
- Each compartment advances to the next stage of development by synthesizing a new sigma factor that displaces the previous sigma factor.
- Communication between the two compartments coordinates the timing of sigma factor substitutions.

Perhaps the most extensive example of switches

in sigma factors is provided by **sporulation**, an alternative lifestyle available to some bacteria. At the end of the **vegetative phase** in a bacterial culture, logarithmic growth ceases because nutrients in the medium become depleted. This triggers sporulation, as illustrated in **FIGURE 11.41**. DNA is replicated, a genome is segregated at one end of the cell, and eventually the genome is surrounded by the tough spore coat. When the septum forms, it generates two independent compartments, the mother cell and the forespore. At the start of the process, one chromosome is attached to each pole of the cell. The growing septum traps part of one chromosome in the forespore, and then a translocase (SpoIIE) pumps the rest of the chromosome into the forespore.

Sporulation takes approximately eight

hours. It can be viewed as a primitive sort of

differentiation, in which a parent cell (the vegetative bacterium) gives rise to two different

daughter cells with distinct fates: the mother

Alternative sigmas control phage development

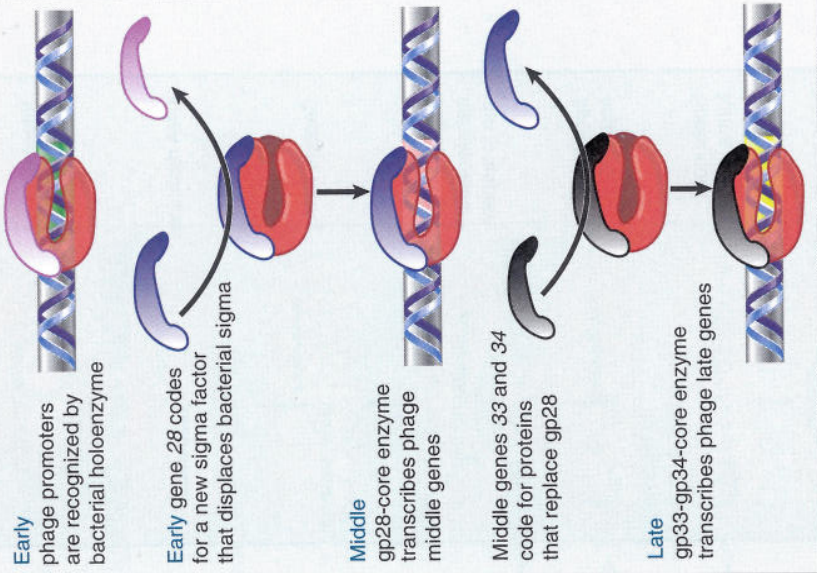


FIGURE 11.40 Transcription of phage SP01 genes is controlled by two successive substitutions of the sigma factor that change the initiation specificity.

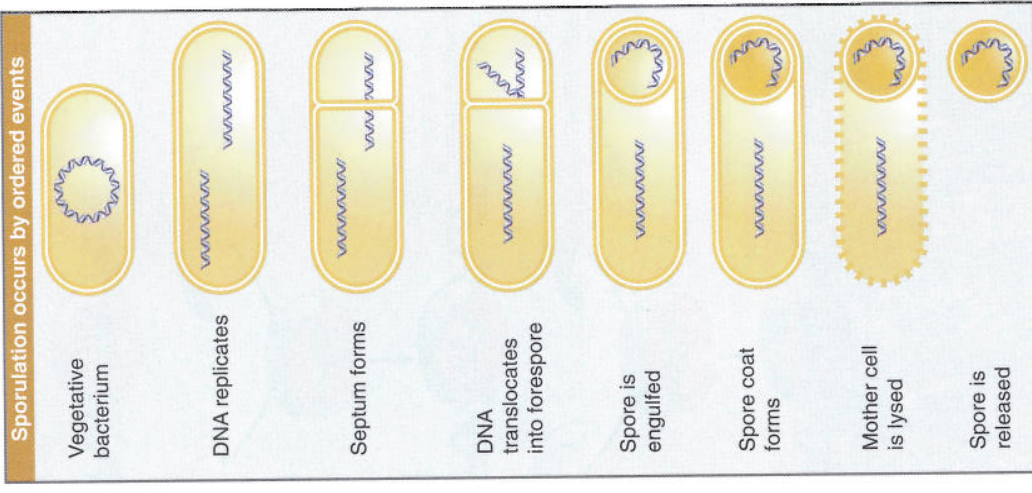


FIGURE 11.41 Sporulation involves the differentiation of a vegetative bacterium into a mother cell that is lysed and a spore that is released.

cell is eventually lysed, and the spore that is released has an entirely different structure from the original bacterium.

Sporulation involves a drastic change in the biosynthetic activities of the bacterium, in which many genes are involved. The basic level of control lies at transcription. Some of the genes that functioned in the vegetative phase are turned off during sporulation, but most continue to be expressed. In addition, the genes specific for sporulation are expressed only during this period. At the end of sporulation, ~40% of the bacterial mRNA is sporulation specific.

New forms of the RNA polymerase become active in sporulating cells; they contain the same core enzyme as vegetative cells, but have different proteins in place of the vegetative σ^{43} . The changes in transcriptional specificity are sum-

marized in **FIGURE 11.42**. The principle is that in each compartment the existing sigma factor is successively displaced by a new factor that causes transcription of a different set of genes. Communication between the compartments occurs in order to coordinate the timing of the changes in the forespore and mother cell.

The sporulation cascade is initiated when environmental conditions trigger a **phosphorelay**, in which a phosphate group is passed along a series of proteins until it reaches SpoOA. (Several gene products are involved in this process, whose complexity may reflect the need to avoid mistakes in triggering sporulation unnecessarily.) SpoOA is a transcriptional regulator whose activity is affected by phosphorylation. In the phosphorylated form, it activates transcription of two operons, each of which is transcribed by a different form of the host RNA polymerase. Under the direction of phosphorylated SpoOA, host enzyme utilizing the general σ^{43} transcribes the gene coding for the factor σ^F , and host enzyme under the direction of a minor factor, σ^H , transcribes the gene coding for the factor pro- σ^E . Both of these new sigma factors are produced before septum formation, but become active later.

Factor σ^F is the first one to become active in the forespore compartment. It is inhibited by an antisigma factor that binds to it; in the forespore, an anti-antisigma factor removes the inhibitor. This reaction is controlled by a series of phosphorylation/dephosphorylation events. The initial determinant is a phosphatase (SpoIIE) that is an integral membrane protein that accumulates at the pole, with the result that its phosphatase domain becomes more concentrated in the forespore. It dephosphorylates, and thereby activates, SpoIIAA, which in turn displaces the antisigma factor SpoIIB from the complex of SpoIIB- σ^F . Release of σ^F activates it.

Activation of σ^F is the start of sporulation. Under the direction of σ^F , RNA polymerase transcribes the first set of sporulation genes instead of the vegetative genes it was previously transcribing. The replacement reaction probably affects only part of the RNA polymerase population, because σ^F is produced only in small amounts. Some vegetative enzyme remains present during sporulation. The displaced σ^{43} is not destroyed; it can be recovered from extracts of sporulating cells.

Two regulatory events follow from the activity of σ^F , as detailed in **FIGURE 11.43**. In the forespore itself, another factor, σ^G , is the product of one of the early sporulation genes. Factor σ^G

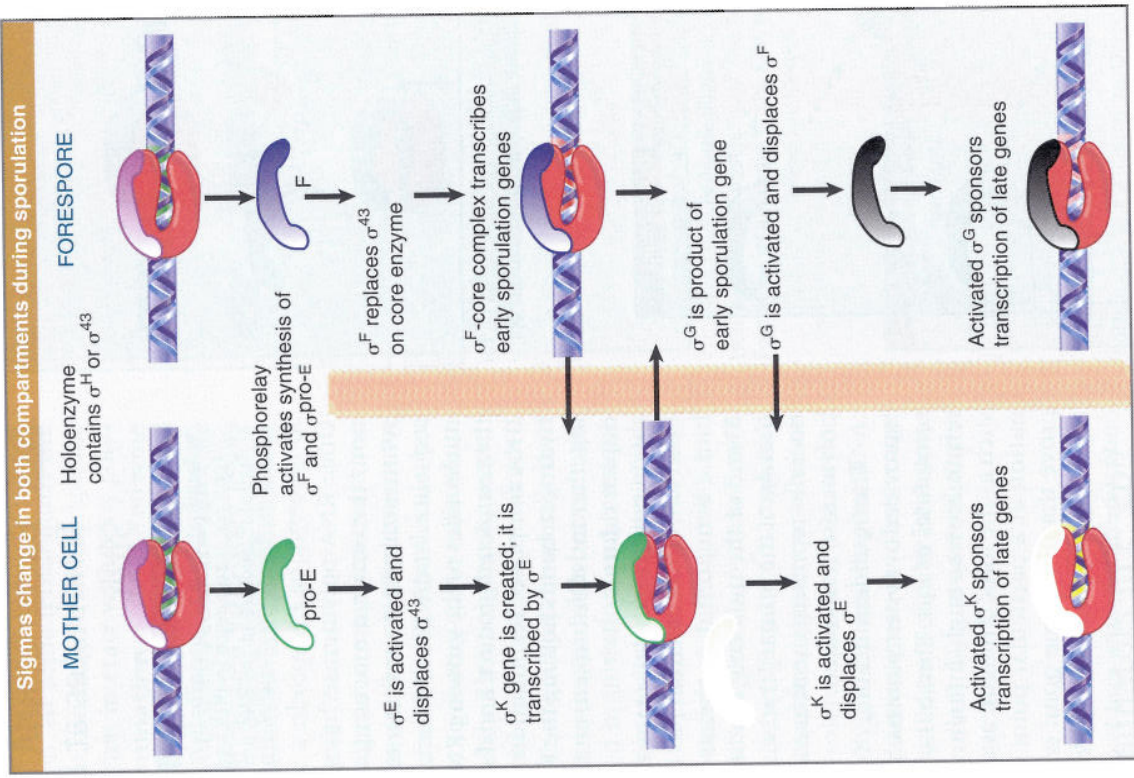


FIGURE 11.42 Sporulation involves successive changes in the sigma factors that control the initiation specificity of RNA polymerase. The cascades in the mother cell (left) and the forespore (right) are related by signals passed across the septum (indicated by horizontal arrows).

causes RNA polymerase to transcribe the late sporulation genes in the forespore. Another early sporulation gene product is responsible for communicating with the mother cell compartment. Factor σ^F activates SpoIIR, which is secreted from the forespore. It then activates the membrane-bound protein SpoIIGA to cleave the inactive precursor pro- σ^E into the active factor σ^E in the mother cell. (Any σ^E that is produced in the forespore is degraded by forespore-specific functions.)

The cascade continues when σ^E in turn is replaced by σ^K . (Actually, the production of σ^K is quite complex, because first its gene must be

created by a recombination event!) This factor also is synthesized as an inactive precursor (pro- σ^K) that is activated by a protease. Once σ^K has been activated, it displaces σ^E and causes transcription of the late genes in the mother cell. The timing of these events in the two compartments are coordinated by further signals. The activity of σ^E in the mother cell is necessary for activation of σ^G in the forespore; in turn the activity of σ^G is required to generate a signal that is transmitted across the septum to activate σ^K .

Sporulation is thus controlled by two cascades, in which sigma factors in each

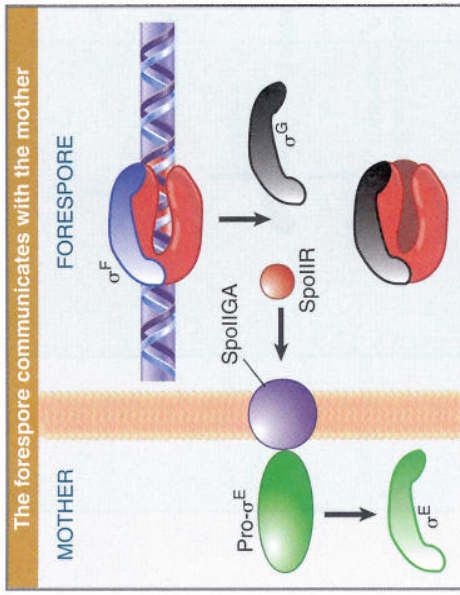


FIGURE 11.43 σ^F triggers synthesis of the next sigma factor in the forespore (σ^E) and turns on SpoIIR, which causes SpoIIA to cleave pro- σ^E .

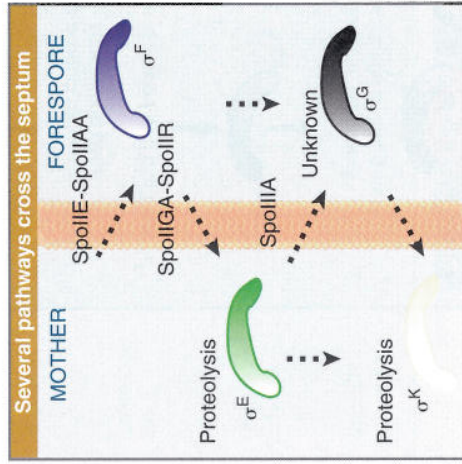


FIGURE 11.44 The crisscross regulation of sporulation coordinates timing of events in the mother cell and forespore.

compartment are successively activated, each directing the synthesis of a particular set of genes. **FIGURE 11.44** outlines how the two cascades are connected by the transmission of signals from one compartment to the other. As new sigma factors become active, old sigma factors are displaced, so that transitions in sigma factors turn genes off as well as on. The incorporation of each factor into RNA polymerase dictates when its set of target genes is expressed, and the amount of factor available influences the level of gene expression. More than one sigma factor may be active at any time, and the specificities of some of the sigma factors overlap. We do not know what is responsible for the ability of each sigma factor to replace its predecessor.

11.20 Bacterial RNA Polymerase Terminates at Discrete Sites

Key concept

- Termination may require both recognition of the terminator sequence in DNA and the formation of a hairpin structure in the RNA product.

Once RNA polymerase has started transcription, the enzyme moves along the template, synthesizing RNA, until it meets a **terminator** sequence. At this point, the enzyme stops adding nucleotides to the growing RNA chain, releases the completed product, and dissociates from the DNA template. Termination requires that all hydrogen bonds holding the RNA–DNA hybrid together must be broken, after which the DNA duplex reforms.

It is difficult to define the termination point of an RNA molecule that has been synthesized in the living cell. It is always possible that the 3' end of the molecule has been generated by cleavage of the primary transcript, and therefore does not represent the actual site at which RNA polymerase terminated.

The best identification of termination sites is provided by systems in which RNA polymerase terminates *in vitro*. The ability of the enzyme to terminate is strongly influenced by parameters such as the ionic strength; as a result, its termination at a particular point *in vitro* does not prove that this same point is a natural terminator. We can, however, identify authentic 3' ends when the same end is generated *in vitro* and *in vivo*.

FIGURE 11.45 summarizes the two types of features found in bacterial terminators.

- Terminators in bacteria and their phages have been identified as sequences that are needed for the termination reaction (*in vitro* or *in vivo*). The sequences at prokaryotic terminators show no similarities beyond the point at which the last base is added to the RNA. The responsibility for termination lies with the sequences *already transcribed* by RNA polymerase. Thus termination relies on scrutiny of the template or product that the polymerase is currently transcribing.
- Many terminators require a hairpin to form in the secondary structure of the RNA being transcribed. *This indicates that termination depends on the RNA product and is not determined simply by scrutiny of the DNA sequence during transcription.*

Terminators vary widely in their efficiencies of termination. At some terminators, the termination event can be *prevented* by specific ancillary factors that interact with RNA polymerase. **Antitermination** causes the enzyme to continue transcription past the terminator sequence, an event called **readthrough** (the same term used in Section 9.14, Suppressors May Compete with Wild-Type Reading of the Code, to describe a ribosome's suppression of termination codons).

In approaching the termination event, we must regard it not simply as a mechanism for generating the 3' end of the RNA molecule, but as an opportunity to control gene expression. Thus the stages when RNA polymerase associates with DNA (initiation) or dissociates from it (termination) both are subject to specific control. There are interesting parallels between the systems employed in initiation and termination. Both require breaking of hydrogen bonds (initial melting of DNA at initiation and RNA–DNA dissociation at termination), and both require additional proteins to interact with the core enzyme. In fact, they are accomplished by alternative forms of the polymerase. Whereas initiation relies solely upon the interaction between RNA polymerase and duplex DNA, though, the termination event involves recognition of signals in the transcript by RNA polymerase or by ancillary factors as well as the recognition of sequences in DNA.

11.21 There Are Two Types of Terminators in *E. coli*

Key concept

- Intrinsic terminators consist of a G–C-rich hairpin in the RNA product followed by a U-rich region in which termination occurs.

Terminators are distinguished in *E. coli* according to whether RNA polymerase requires any additional factors to terminate *in vitro*:

- Core enzyme can terminate *in vitro* at certain sites in the absence of any other factor. These sites are called **intrinsic terminators**.
- Rho-dependent** terminators are defined by the need for addition of **rho factor** (ρ) *in vitro*, and mutations show that the factor is involved in termination *in vivo*.

Intrinsic terminators have the two structural features evident in **FIGURE 11.46**: a hairpin in the secondary structure, and a region at the

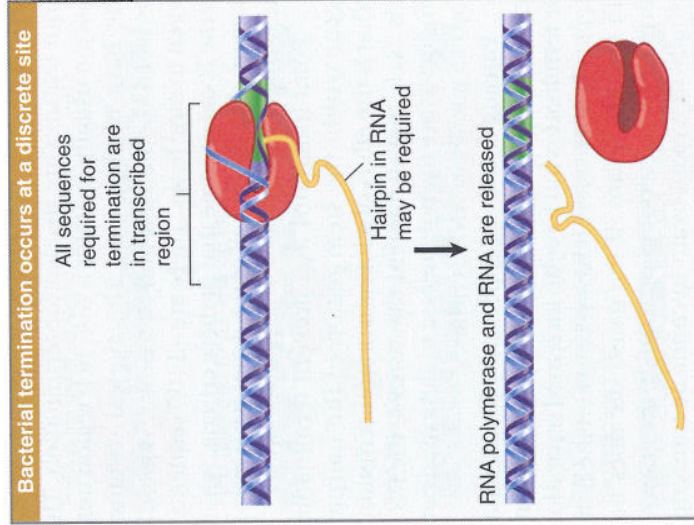


FIGURE 11.45 The DNA sequences required for termination are located upstream of the terminator sequence. Formation of a hairpin in the RNA may be necessary.

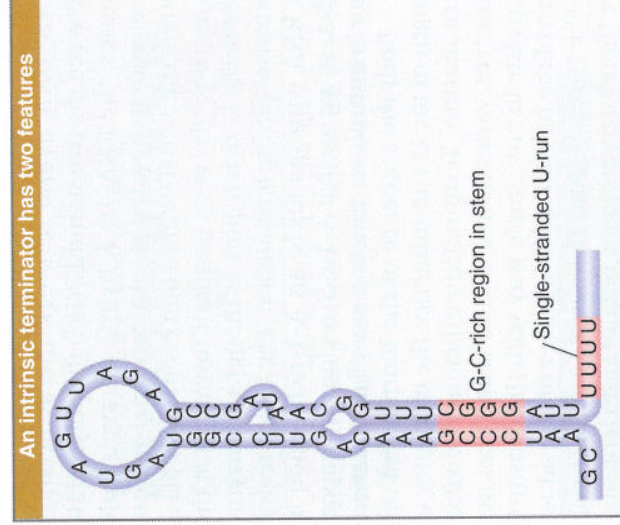


FIGURE 11.46 Intrinsic terminators include palindromic regions that form hairpins varying in length from 7 to 20 bp. The stem-loop structure includes a G–C-rich region and is followed by a run of U residues.

11.22 How Does Rho Factor Work?

Key concept

- Rho factor is a terminator protein that binds to a *rut* site on nascent RNA and tracks along the RNA to release it from the RNA-DNA hybrid structure at the RNA polymerase.

very end of the unit that is rich in U residues. Both features are needed for termination. The hairpin usually contains a G-C-rich region near the base of the stem. The typical distance between the hairpin and the U-rich region is seven to nine bases. There are ~1100 sequences in the *E. coli* genome that fit these criteria, which suggests that about half of the genes have intrinsic terminators.

Point mutations that prevent termination occur within the stem region of the hairpin. What is the effect of a hairpin on transcription? It is likely that all hairpins that form in the RNA product cause the polymerase to slow (and perhaps pause) in RNA synthesis.

Pausing creates an opportunity for termination to occur. Pausing occurs at sites that resemble terminators but have an increased separation (typically ten to eleven bases) between the hairpin and the U-run. If the pause site does not correspond to a terminator, though, the enzyme usually moves on again to continue transcription. The length of the pause varies, but at a typical terminator lasts ~60 seconds.

A downstream U-rich region destabilizes the RNA-DNA hybrid when RNA polymerase pauses at the hairpin. The rU-dA RNA-DNA hybrid has an unusually weak base-paired structure; it requires the least energy of any RNA-DNA hybrid to break the association between the two strands. When the polymerase pauses, the RNA-DNA hybrid unravels from the weakly bonded rU-dA terminal region. Often, the actual termination event takes place at any one of several positions toward or at the end of the U-rich region, as though the enzyme “stutters” during termination. The U-rich region in RNA corresponds to an A-T-rich region in DNA, so we see that A-T-rich regions are important in intrinsic termination as well as initiation.

Both the sequence of the hairpin and the length of the U-run influence the efficiency of termination. Termination efficiency *in vitro*, however, varies from 2% to 90% and does not correlate in any simple way with the constitution of the hairpin or the number of U residues in the U-rich region. The hairpin and U-region are therefore necessary, but not sufficient, and additional parameters influence the interaction with RNA polymerase. In particular, the sequences both upstream and downstream of the intrinsic terminator influence its efficiency.

Less is known about the signals and ancillary factors involved in termination for eukaryotic polymerases. Each class of polymerase uses a different mechanism (see Chapter 26, RNA Splicing and Processing).

domains, and pushes the 5' end of the bound region into the interior, where it is bound by a secondary RNA-binding domain in the C-terminal domains. The initial form of rho is a gapped ring, but binding of the RNA converts it to a closed ring.

After binding to the *rut* site, rho uses its helicase activity, driven by ATP hydrolysis, to translocate along RNA until it reaches the RNA-DNA hybrid stretch in RNA polymerase. It then unwinds the duplex structure. We do not know whether this action is sufficient to release the transcript or whether rho also interacts with RNA polymerase to help release RNA. Rho functions as an ancillary factor for RNA polymerase; typically its maximum activity *in vitro* is displayed when it is present at ~10% of the concentration of the RNA polymerase.

Rho needs to translocate along RNA from the *rut* site to the actual point of termination. This requires the factor to move faster than RNA polymerase. The enzyme pauses when it reaches a terminator, and termination occurs if rho catches it there. Pausing is therefore important in rho-dependent termination, just as in intrinsic termination, because it gives time for the other necessary events to occur.

The idea that rho moves along RNA leads to an important prediction about the relationship between transcription and translation. Rho must first have access to a binding sequence on RNA and then must be able to move along the RNA. Either or both of these conditions may be prevented if ribosomes are translating an RNA. Thus the ability of rho factor to reach RNA polymerase at a terminator depends on what is happening in translation.

This model explains a puzzling phenomenon. In some cases, a nonsense mutation in one gene of a transcription unit prevents the expression of subsequent genes in the unit. This effect is called **polarity**. A common cause is the

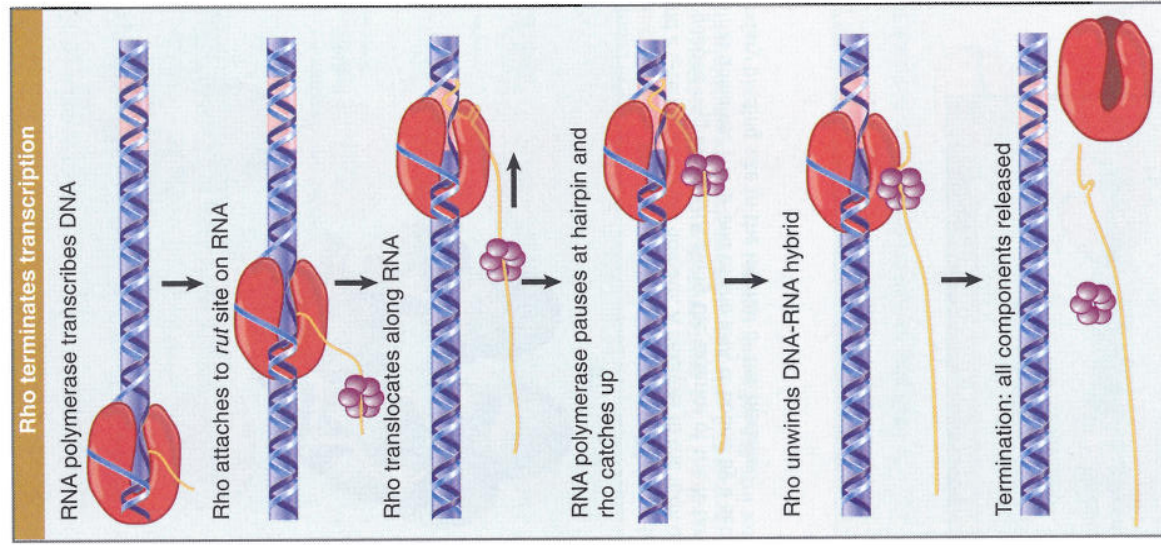


FIGURE 11.47 Rho factor binds to RNA at a *rut* site and translocates along RNA until it reaches the RNA-DNA hybrid in RNA polymerase, where it releases the RNA from the DNA.

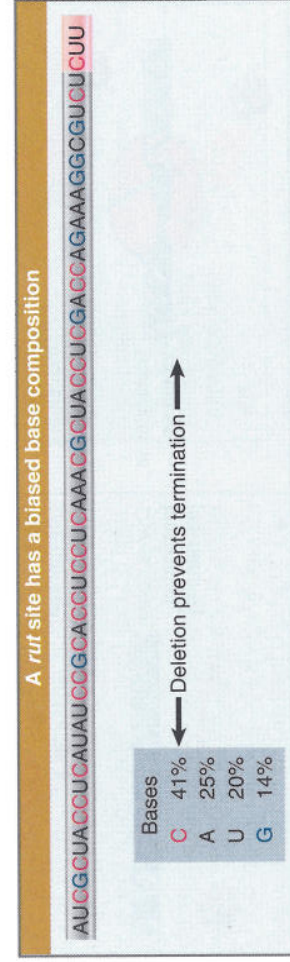


FIGURE 11.48 A *rut* site has a sequence rich in C and poor in G preceding the actual site(s) of termination. The sequence corresponds to the 3' end of the RNA.

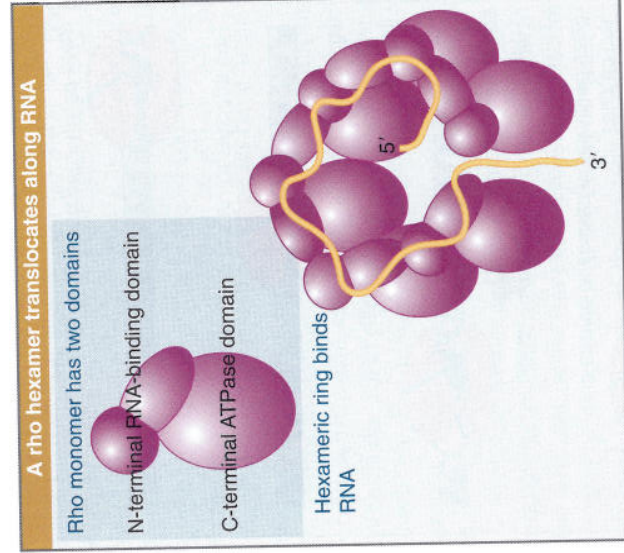


FIGURE 11.49 Rho has an N-terminal RNA-binding domain and a C-terminal ATPase domain. A hexamer in the form of a gapped ring binds RNA along the exterior of the N-terminal domains. The 5' end of the RNA is bound by a secondary binding site in the interior of the hexamer.

absence of the mRNA corresponding to the subsequent (distal) parts of the unit.

Suppose that there are rho-dependent terminators *within* the transcription unit, that is, before the terminator that *usually* is used. The consequences are illustrated in **FIGURE 11.50**. Normally these earlier terminators are not used, because the ribosomes prevent rho from reaching RNA polymerase. A nonsense mutation, however, releases the ribosomes, so that rho is free to attach to and/or move along the mRNA, enabling it to act on RNA polymerase at the terminator. As a result, the enzyme is released, and the distal regions of the transcription unit are never expressed. (Why should there be internal terminators? Perhaps they are simply sequences that by coincidence mimic the usual rho-dependent terminator.) Some stable RNAs that have extensive secondary structure are preserved from polar effects, presumably because the structure impedes rho attachment or movement.

rho mutations show wide variations in their influence on termination. The basic nature of the effect is a failure to terminate. The magnitude of the failure, however, as seen in the percent of readthrough *in vivo*, depends on the particular target locus. Similarly, the need for rho

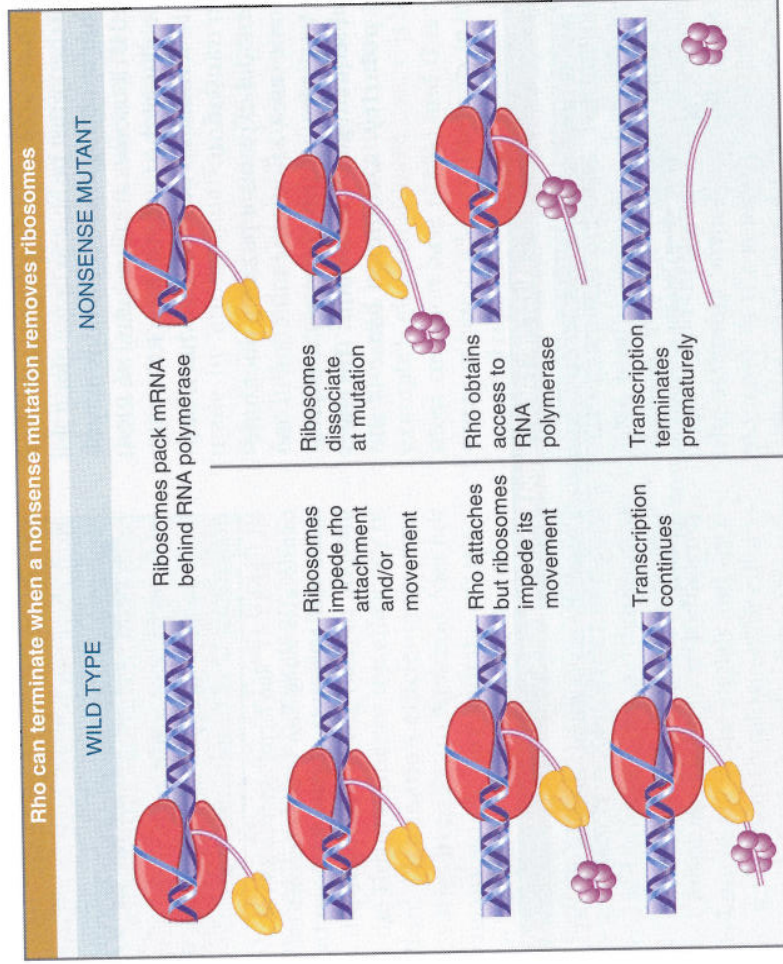


FIGURE 11.50 The action of rho factor may create a link between transcription and translation when a rho-dependent terminator lies soon after a nonsense mutation.

factor *in vitro* is variable. Some (rho-dependent) terminators require relatively high concentrations of rho, whereas others function just as well at lower levels. This suggests that different terminators require different levels of rho factor for termination and therefore respond differently to the residual levels of rho factor in the mutants (*rho* mutants are usually leaky).

Some *rho* mutations can be suppressed by mutations in other genes. This approach provides an excellent way to identify proteins that interact with rho. The β subunit of RNA polymerase is implicated by two types of mutation. First, mutations in the *rpoB* gene can reduce termination at a rho-dependent site. Second, mutations in *rpoB* can restore the ability to terminate transcription at rho-dependent sites in *rho* mutant bacteria. We do not, however, know what function the interaction plays.

11.23 Antitermination Is a Regulatory Event

Key concepts

- Termination is prevented when antitermination proteins act on RNA polymerase to cause it to read through a specific terminator or terminators.
- Phage lambda has two antitermination proteins, pN and pQ, that act on different transcription units.

Antitermination is used as a control mechanism in both phage regulatory circuits and bacterial operons. **FIGURE 11.51** shows that antitermination controls the ability of the enzyme to read past a terminator into genes lying beyond. In the example shown in the figure, the default pathway is for RNA polymerase to terminate at the end of region 1; however, antitermination allows it to continue transcription through region 2. The promoter does not change, so as a result both situations produce an RNA with the same 5' sequences; the difference is that after antitermination the RNA is extended to include new sequences at the 3' end.

Antitermination was discovered in bacteriophage infections. A common feature in the control of phage infection is that very few of the phage genes (the "early" genes) can be transcribed by the bacterial host RNA polymerase. Among these genes, however, are regulator(s) whose product(s) allow the next set of phage genes to be expressed (see Section 14.4, Two Types of Regulatory Event Control the Lytic Cascade). One of these types of regulator is an **antitermination protein**. **FIGURE 11.52** shows

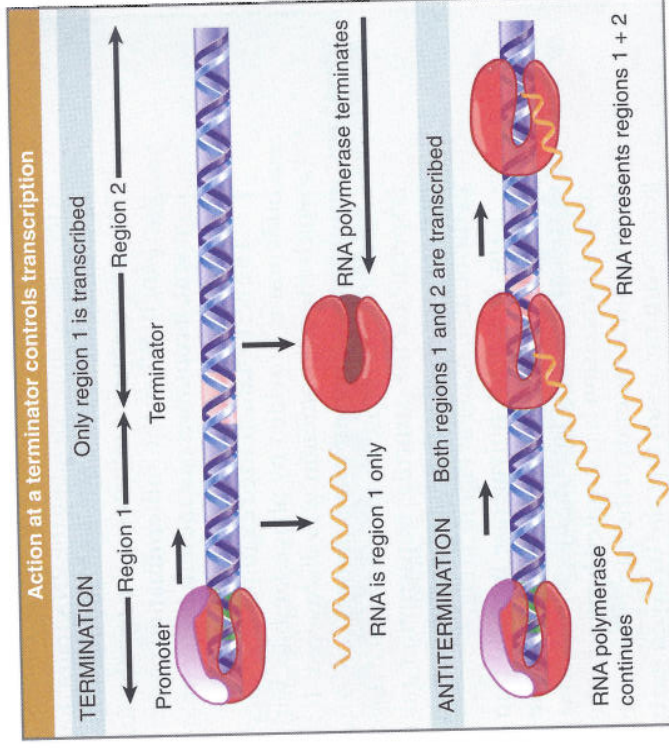


FIGURE 11.51 Antitermination can control transcription by determining whether RNA polymerase terminates or reads through a particular terminator into the following region.

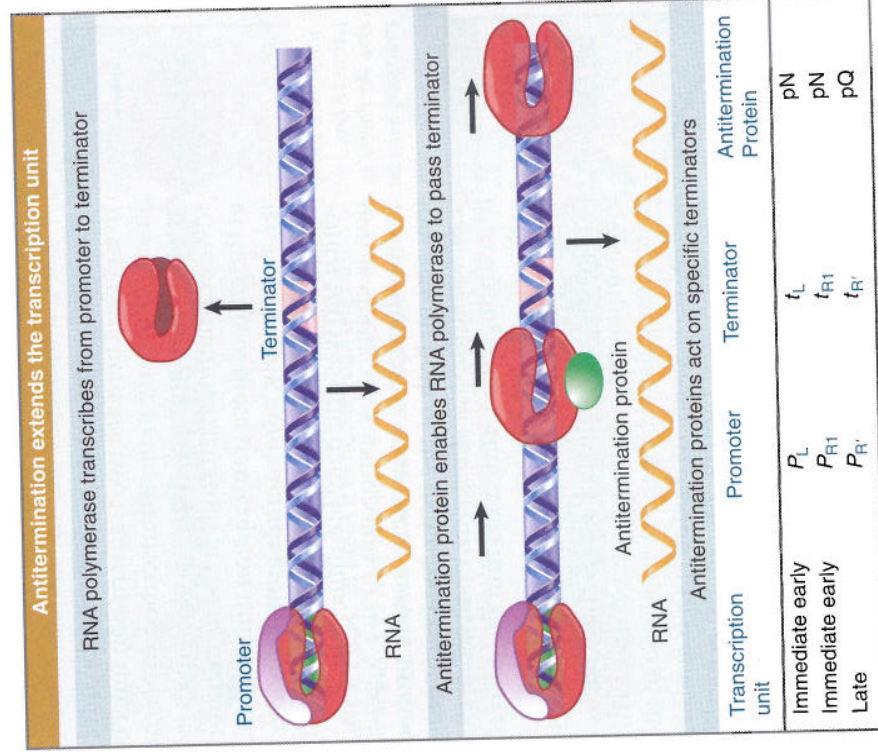


FIGURE 11.52 An antitermination protein can act on RNA polymerase to enable it to read through a specific terminator.

that it enables RNA polymerase to read through a terminator, thus extending the RNA transcript. In the absence of the antitermination protein, RNA polymerase terminates at the terminator (top panel). When the antitermination protein is present, it continues past the terminator (middle panel).

The best-characterized example of antitermination is provided by phage lambda, with which the phenomenon was discovered. It is used at two stages of phage expression. The antitermination protein produced at each stage is specific for the particular transcription units that are expressed at that stage, as summarized in the bottom panel of Figure 11.52.

The host RNA polymerase initially transcribes two genes, which are called the **immediate early** genes. The transition to the next stage of expression is controlled by preventing termination at the ends of the immediate early genes, with the result that the **delayed early** genes are expressed. (We discuss the overall regulation of lambda development in Chapter 14, Phage Strategies.)

The regulator gene that controls the switch from immediate early to delayed early expression is identified by mutations in lambda gene *N* that can transcribe *only* the immediate early genes; they proceed no further into the infective cycle. There are two transcription units of immediate early genes (transcribed from the promoters P_L and P_R). Transcription by *E. coli* RNA polymerase itself stops at the terminators at the ends of these transcription units (t_{L1} and t_{R1} , respectively.) Both terminators depend on rho; in fact, these were the terminators with which rho was originally identified. The situation is changed by expression of the *N* gene. The product pN is an antitermination protein that acts on both of the immediate early transcription units and allows RNA polymerase to read through the terminators into the delayed early genes beyond them.

As for other phages, still another control is needed to express the late genes that code for the components of the phage particle. This switch is regulated by gene *Q*, itself one of the delayed early genes. Its product, pQ, is another antitermination protein, one that specifically allows RNA polymerase initiating at another site, the late promoter $P_{R'}$, to read through a terminator that lies between it and the late genes.

The different specificities of pN and pQ establish an important general principle: *RNA polymerase interacts with transcription units in such a*

way that an ancillary factor can sponsor antitermination specifically for some transcripts. Termination can be controlled with the same sort of precision as initiation.

11.24 Antitermination Requires Sites That Are Independent of the Terminators

Key concepts

- The site where an antitermination protein acts is upstream of the terminator site in the transcription unit.
- The location of the antitermination site varies in different cases and can be in the promoter or within the transcription unit.

Which sites are involved in controlling the specificity of antitermination? The antitermination activity of pN is highly specific, but the *antitermination event is not determined by the terminators t_{L1} and t_{R1} ; the recognition site needed for antitermination lies upstream in the transcription unit, that is, at a different place from the terminator site at which the action eventually is accomplished.* **FIGURE 11.53** shows the locations of the sites required for antitermination in phage lambda.

The recognition sites required for pN action are called *nut* (for *N* utilization). The sites responsible for determining leftward and rightward antitermination are described as *nutL* and *nutR*, respectively. Mapping of *nut* mutations locates *nutL* between the startpoint of P_L and the beginning of the *N* coding region. By contrast, *nutR* lies between the end of the *cro* gene and t_{L1} . This means that the two *nut* sites lie in different positions relative to the organization of their transcription units. Whereas *nutL* is near the promoter, *nutR* is near to the terminator. (*nut* is different yet again, and lies within the promoter.)

How does antitermination occur? When pN recognizes the *nut* site, it must act on RNA polymerase to ensure that the enzyme can no longer respond to the terminator. The variable locations of the *nut* sites indicate that this event is linked neither to initiation nor to termination, but can occur to RNA polymerase as it elongates the RNA chain past the *nut* site. As illustrated in **FIGURE 11.54**, the polymerase then becomes a juggernaut that continues past the terminator, heedless of its signal. (This reaction involves antitermination at rho-dependent ter-

minators, but pN also suppresses termination at intrinsic terminators.)

Is the ability of pN to recognize a short sequence within the transcription unit an example of a more widely used mechanism for antitermination? Phages that are related to lambda have different *N* genes and different antitermination specificities. The region of the phage genome in which the *nut* sites lie has a different sequence in each of these phages, and each phage must therefore have characteristic *nut* sites recognized specifically by its own pN. Each of these pN products must have the same general ability to interact with the transcription apparatus in an antitermination capacity, but also have a different specificity for the sequence of DNA that activates the mechanism.

11.25 Termination and Antitermination Factors Interact with RNA Polymerase

Key concepts

- Several bacterial proteins are required for lambda pN to interact with RNA polymerase.
- These proteins are also involved in antitermination in the *rrn* operons of the host bacterium.
- The lambda antitermination pQ has a different mode of interaction that involves binding to DNA at the promoter.

Termination and antitermination are closely connected and involve bacterial and phage proteins that interact with RNA polymerase. Several proteins concerned with termination have been identified by isolating mutants of *E. coli* in which pN is ineffective. Several of these mutations lie in the *rpoB* gene. This argues that pN (like rho factor) interacts with the β subunit of the core enzyme. Other *E. coli* mutations that prevent pN function identify the *nus* loci: *nusA*, *nusB*, *nusE*, and *nusG*. (The term “*nus*” is an acronym for *N* utilization substance.)

A lambda *nut* site consists of two sequence elements called *boxA* and *boxB*. Sequence elements related to *boxA* are also found in bacterial operons. *boxA* is required for binding bacterial proteins that are necessary for antitermination in both phage and bacterial operons. *boxB* is specific to the phage genome, and mutations in *boxB* abolish the ability of pN to cause antitermination.

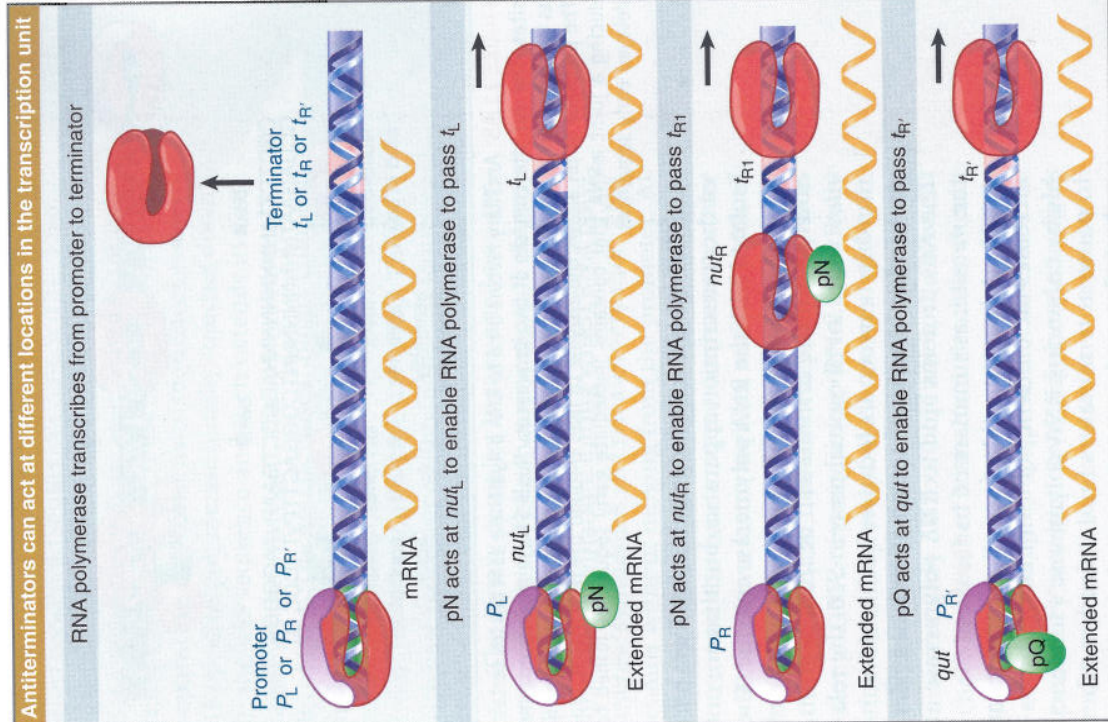


FIGURE 11.53 Host RNA polymerase transcribes lambda genes and terminates at *t* sites. pN allows it to read through terminators in the *L* and *R1* units; pQ allows it to read through the *R'* terminator. The sites at which pN acts (*nut*) and at which pQ acts (*qut*) are located at different relative positions in the transcription units.

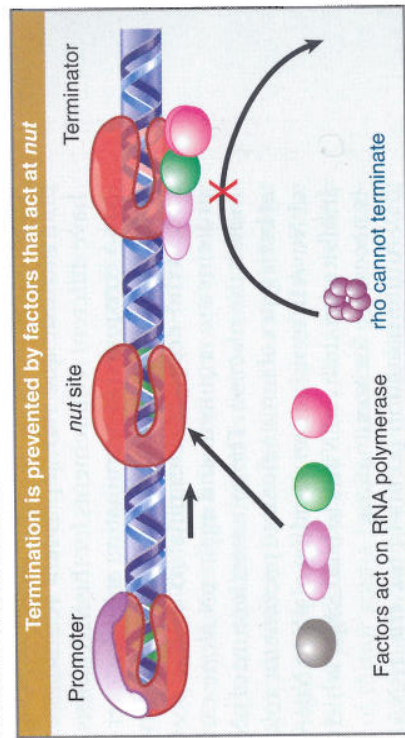


FIGURE 11.54 Ancillary factors bind to RNA polymerase as it passes the *nut* site. They prevent rho from causing termination when the polymerase reaches the terminator.

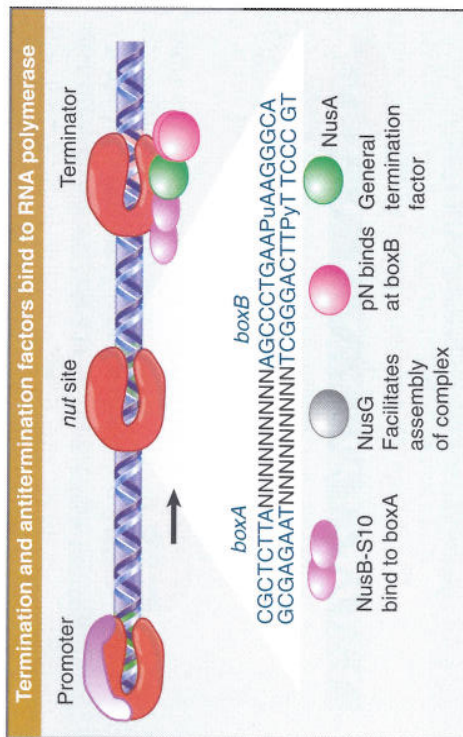


FIGURE 11.55 Ancillary factors bind to RNA polymerase as it passes certain sites. The *nut* site consists of two sequences. NusB-S10 joins core enzyme as it passes *boxA*. Then NusA and pN protein bind as polymerase passes *boxB*. The presence of pN allows the enzyme to read through the terminator, producing a joint mRNA that contains immediate early sequences joined to delayed early sequences.

The *nus* loci code for proteins that form part of the transcription apparatus, but that are not isolated with the RNA polymerase enzyme. The *nusA*, *nusB*, and *nusG* functions are concerned solely with the termination of transcription. *nusE* codes for ribosomal protein S10; the relationship between its location in the 30S subunit and its function in termination is not clear. The Nus proteins bind to RNA polymerase at the *nut* site, as summarized in **FIGURE 11.55**.

NusA is a general transcription factor that increases the efficiency of termination, probably by enhancing RNA polymerase's tendency to pause at terminators (and indeed at other regions of secondary structure; see below). NusB and S10 form a dimer that binds specifically to RNA containing a *boxA* sequence. NusG may be concerned with the general assembly of all the Nus factors into a complex with RNA polymerase.

Intrinsic and rho-dependent terminators have different requirements for the Nus factors. NusA is required for termination at intrinsic terminators, and the reaction can be prevented by pN. At rho-dependent terminators, all four Nus proteins are required, and again pN alone can inhibit the reaction. The common feature of pN at both types of terminator is to prevent the role of NusA in termination. Binding of pN to NusA inhibits the ability of NusA to bind RNA, which is necessary for termination.

Antitermination occurs in the *rrn* (rRNA) operons of *E. coli* and involves the same *nus* functions. The leader regions of the *rrn* operons contain *boxA* sequences; NusB-S10 dimers rec-

ognize these sequences and bind to RNA polymerase as it elongates past *boxA*. This changes the properties of RNA polymerase in such a way that it can now read through rho-dependent terminators that are present within the transcription unit.

The *boxA* sequence of lambda RNA does not bind NusB-S10 and is probably enabled to do so by the presence of NusA and pN; the *boxB* sequence could be required to stabilize the reaction. Thus variations in *boxA* sequences may determine which particular set of factors is required for antitermination. The consequences are the same: when RNA polymerase passes the *nut* site, it is modified by addition of appropriate factors and fails to terminate when it subsequently encounters the terminator sites.

Antitermination in lambda requires pN to bind to RNA polymerase in a manner that depends on the sequence of the transcription unit. Does pN recognize the *boxB* site in DNA or in the RNA transcript? It does not bind independently to either type of sequence, but does bind to a transcription complex when core enzyme passes the *boxB* site. pN has separate domains that recognize the *boxB* RNA sequence and the NusA protein. After joining the transcription complex, pN remains associated with the core enzyme, in effect becoming an additional subunit whose presence changes recognition of terminators. It is possible that pN in fact continues to bind to both the *boxB* RNA sequence and to RNA polymerase, which maintains a loop in the RNA; thus the role of *boxB* RNA would partly be to tether pN in the vicinity, effectively increasing its local concentration.

pQ, which prevents termination later in phage infection by acting at *qut*, has a different mode of action. The *qut* sequence lies at the start of the late transcription unit. The upstream part of *qut* lies within the promoter, whereas the downstream part lies at the beginning of the transcribed region. This implies that pQ action involves recognition of DNA; it also implies that its mechanism of action, at least concerning the initial binding to the complex, must be different from that of pN. pQ interacts with the holoenzyme during the initiation phase. In fact, σ^{70} is required for the interaction with pQ. This reinforces the view of RNA polymerase as an interactive structure in which conformational changes induced at one phase may affect its activity at a later phase.

The basic action of pQ is to interfere with pausing. Once pQ has acted upon RNA polymerase, the enzyme shows much reduced paus-

ing at all sites, including rho-dependent and intrinsic terminators. This means that pQ does not act directly on termination *per se*, but instead allows the enzyme to pass the terminator more quickly, thus depriving the core polymerase and/or accessory factor of the opportunity to cause termination.

The general principle is that RNA polymerase may exist in forms that are competent to undertake particular stages of transcription, and its activities at these stages can be changed only by modifying the appropriate form. Thus substitutions of sigma factors may change one initiation-competent form into another; and additions of Nus factors may change the properties of termination-competent forms.

Termination seems to be closely connected with the mode of elongation. In its basic transcription mode, core polymerase is subject to many pauses during elongation, and pausing at a terminator site is the prerequisite for termination to occur. Under the influence of factors such as NusA, pausing becomes extended, increasing the efficiency of termination; while under the influence of pN or pQ, pausing is abbreviated, decreasing the efficiency of termination. Recognition sites for these factors are found only in certain transcription units, and as a result pausing and consequently termination are altered only in those units.

11.26 Summary

A transcription unit comprises the DNA between a promoter, where transcription initiates, and a terminator, where it ends. One strand of the DNA in this region serves as a template for synthesis of a complementary strand of RNA. The RNA-DNA hybrid region is short and transient, as the transcription "bubble" moves along DNA. The RNA polymerase holoenzyme that synthesizes bacterial RNA can be separated into two components. Core enzyme is a multimer of structure $\alpha_2\beta\beta'$ that is responsible for elongating the RNA chain. Sigma factor (σ) is a single subunit that is required at the stage of initiation for recognizing the promoter.

Core enzyme has a general affinity for DNA. The addition of sigma factor reduces the affinity of the enzyme for nonspecific binding to DNA, but increases its affinity for promoters. The rate at which RNA polymerase finds its promoters is too great to be accounted for by diffusion and random contacts with DNA; direct exchange of DNA sequences held by the enzyme may be involved.

Bacterial promoters are identified by two short conserved sequences centered at -35 and -10 relative to the startpoint. Most promoters have sequences that are well related to the consensus sequences at these sites. The distance separating the consensus sequences is 16 to 18 bp. RNA polymerase initially "touches down" at the -35 sequence and then extends its contacts over the -10 region. The enzyme covers ~77 bp of DNA. The initial "closed" binary complex is converted to an "open" binary complex by melting of a sequence of ~12 bp that extends from the -10 region to the startpoint. The A-T-rich base pair composition of the -10 sequence may be important for the melting reaction.

The binary complex is converted to a ternary complex by the incorporation of ribonucleotide precursors. There are multiple cycles of abortive initiation, during which RNA polymerase synthesizes and releases very short RNA chains without moving from the promoter. At the end of this stage, there is a change in structure, and the core enzyme contracts to cover ~50 bp. Sigma factor is either released (30% of cases) or changes its form of association with the core enzyme. The core enzyme then moves along DNA, synthesizing RNA. A locally unwound region of DNA moves with the enzyme. The enzyme contracts further in size to cover only 30 to 40 bp when the nascent chain has reached 15 to 20 nucleotides; then it continues to the end of the transcription unit.

The "strength" of a promoter describes the frequency at which RNA polymerase initiates transcription; it is related to the closeness with which its -35 and -10 sequences conform to the ideal consensus sequences, but is influenced also by the sequences immediately downstream of the startpoint. Negative supercoiling increases the strength of certain promoters. Transcription generates positive supercoils ahead of RNA polymerase and leaves negative supercoils behind the enzyme.

The core enzyme can be directed to recognize promoters with different consensus sequences by alternative sigma factors. In *E. coli*, these sigma factors are activated by adverse conditions, such as heat shock or nitrogen starvation. *B. subtilis* contains a single major sigma factor with the same specificity as the *E. coli* sigma factor and also contains a variety of minor sigma factors. Another series of factors is activated when sporulation is initiated; sporulation is regulated by two cascades in which sigma factor replacements occur in the forespore and mother cell. A cascade for

References

- 11.2** Transcription Occurs by Base Pairing in a "Bubble" of Unpaired DNA
Review
Losick, R. and Chamberlin, M. (1976). RNA Polymerase. *Cold Spring Harbor Symp. Quant. Biol.*
- Research*
Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). A structural model of transcription elongation. *Science* 289, 619–625.
- 11.3** The Transcription Reaction Has Three Stages
Research
Rice, G. A., Kane, C. M., and Chamberlin, M. (1991). Footprinting analysis of mammalian RNA polymerase II along its transcript: an alternative view of transcription elongation. *Proc. Natl. Acad. Sci. USA* 88, 4245–4281.
- Wang, D. et al. (1995). Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell* 81, 341–350.
- 11.4** Phase T7 RNA Polymerase Is a Useful Model System
Research
Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399, 80–83.
- Cheetham, G. M. T. and Steitz, T. A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286, 2305–2309.
- Temliakov, D., Mentesana, D., Temliakov, D., Ma, K., Mustaev, A., Borukhov, S., and McAllister, W. T. (2000). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl. Acad. Sci. USA* 97, 14109–14114.
- 11.5** A Model for Enzyme Movement Is Suggested by the Crystal Structure
Review
Shilatifard, A., Conaway, R. C., and Conaway, J. W. (2003). The RNA polymerase II elongation complex. *Annu. Rev. Biochem.* 72, 693–715.
- Research*
Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000). Architecture of RNA polymerase II and implications for

- the transcription mechanism. *Science* 288, 640–649.
- Cramer, P., Bushnell, P., and Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292, 1863–1876.
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292, 1876–1882.
- 11.6** Bacterial RNA Polymerase Consists of Multiple Subunits
Review
Helmann, J. D. and Chamberlin, M. (1988). Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* 57, 839–872.
- Research*
Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104, 901–912.
- Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). A structural model of transcription elongation. *Science* 289, 619–625.
- Zhang, G., Campbell, E. A., Zhang, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811–824.
- 11.7** RNA Polymerase Consists of the Core Enzyme and Sigma Factor
Research
Travers, A. A. and Burgess, R. R. (1969). Cyclic reuse of the RNA polymerase sigma factor. *Nature* 222, 537–540.
- 11.8** The Association with Sigma Factor Changes at Initiation
Research
Bar-Nahum, G. and Nudler, E. (2001). Isolation and characterization of sigma(70)-retaining transcription elongation complexes from *E. coli*. *Cell* 106, 443–451.
- Krummel, B. and Chamberlin, M. J. (1989). RNA chain initiation by *E. coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry* 28, 7829–7842.
- Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonja, E., Ebricht, Y. W., and Ebricht, R. H. (2001). Translocation of sigma(70) with RNA polymerase during transcription. Fluores-
- cence resonance energy transfer assay for movement relative to DNA. *Cell* 106, 453–463.
- 11.9** A Stalled RNA Polymerase Can Restart
Research
Kettenberger, H., Armache, K. J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114, 347–357.
- Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., and Darst, S. A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* 114, 335–345.
- 11.11** Sigma Factor Controls Binding to DNA
Research
Bar-Nahum, G. and Nudler, E. (2001). Isolation and characterization of sigma(70)-retaining transcription elongation complexes from *E. coli*. *Cell* 106, 443–451.
- Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonja, E., Ebricht, Y. W., and Ebricht, R. H. (2001). Translocation of sigma(70) with RNA polymerase during transcription. Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* 106, 453–463.
- 11.12** Promoter Recognition Depends On Consensus Sequences
Review
McClure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* 54, 171–204.
- Research*
Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262, 1407–1413.
- 11.13** Promoter Efficiencies Can Be Increased or Decreased by Mutation
Review
McClure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* 54, 171–204.
- 11.14** RNA Polymerase Binds to One Face of DNA
Review
Siebenlist, U., Simpson, R. B., and Gilbert, W. (1980). *E. coli* RNA polymerase interacts homologically with two different promoters. *Cell* 20, 269–281.